Non-arbitrariness in Novel Sign Systems

Carrie Ann Theisen

Doctor of Philosophy

Institute for Communicating and Collaborative Systems
School of Informatics
University of Edinburgh

2009

Abstract

This thesis investigates non-arbitrariness in novel sets of signs (mappings between signals and meanings). Two common ways of characterizing signs – by the degree of motivatedness they exhibit and by the degree of systematic compositionality they exhibit – are not necessarily orthogonal. Thus, the emergence of arbitrariness and of systematic compositionality in language should be studied together. We focus on a particular interface of the two properties: the systematic re-use of arbitrary elements. Previous work that demonstrates how novel signs can emerge and then evolve to become arbitrary does not measure systematic compositionality. On the other hand, previous work on systematic compositionality proposes a mechanism for the evolution of systematic compositionality and a measure of the property, but does not address the evolution of arbitrariness. We propose a parallel theory of the emergence of the systematic re-use of arbitrary elements. Systematic compositionality emerges in novel, motivated signs and is maintained as the signs become arbitrary.

We report a series of experiments that probe how the systematic re-use of arbitrary elements arises in novel communication systems. In Experiment 1, partners must create signs from scratch to communicate about items that share semantic features. The systematic re-use of arbitrary elements emerges. Further, the evolution of arbitrariness and systematic compositionality are parallel: even participants' first drawings of items are systematically compositional, and this systematic compositionality is maintained as the signs become arbitrary. Experiment 2 demonstrates that naïve participants, who played no role in – indeed, did not even observe – the creation of the sign systems, can nonetheless detect the systematic compositionality in them and generalize from it. Experiment 3 shows that participants actually *do* make use of the systematic re-use of arbitrary elements that they observe in others' sign systems, when faced with the task of communicating (rather than learning and reproducing). The systematic compositionality is not only maintained, but appears to be increasing, over generations of participants observing others' signs. The increase in systematic compositionality occurs when pairs create signs for items they have not observed – presumably, as they generalize using the systematic compositionality they have observed.

In sum, we present an alternative mechanism for the emergence of the systematic re-use of arbitrary elements: arbitrariness and systematic compositionality emerge in a parallel fashion within the dyad, and subsequent communicators maintain – or even increase – the structure they have observed. More generally, we demonstrate the importance of examining arbitrariness and systematic compositionality together.

Acknowledgements

Many thanks to my two supervisors, Jon Oberlander and Simon Kirby. Simon has pushed me to do things I didn't think I was ready for, has invested much time into keeping me connected to the LEC, and has delivered straight talk when necessary. Jon has several times had to act as more of a psychological counsellor than academic advisor, has always found time for me in a pinch, and has indulged me in conversations about theory when I bored of practice. I hope they both look back on supervising me remotely as a risk that paid off, rather than a mistake never to be repeated!

Thanks also to everyone at the LEC for their feedback and practical help on my visits to Edinburgh. It's been so encouraging to know that such intelligent, interesting people care about my work.

On the academic side, I can't forget to thank Eddie Dubourg and Mike Bennett for computing support, and especially Nik Swoboda for making many modifications to Pigeon.

My family and friends have been incredibly supportive over these years. They've never made me feel guilty about being busy and, despite not understanding what I do (or why), have offered to help in any way they could. When I'd get discouraged, I'd pretend I really was as clever and hard working as they seem to think. While many of my family and friends have tested things for me, Jes and John deserve a special mention for their hours and hours of work.

Finally, huge thanks to my boyfriend, Jeremiah. He's supported me financially for years now, but also tolerated my twisted justifications for this. Also, working remotely meant that, most days, he was the only one I could talk to about my work – I know it got boring after a while. Finally, he's been very patient as I've put other parts of our life on hold.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(Carrie Ann Theisen)

Table of Contents

Chapter 1

Arbitrariness and systematic compositionality in language

This thesis explores how a human language can come to map signals to meanings in the unique way that it does. To that end, this introductory chapter lays out the key theoretical concepts. In particular, it:

- makes some preliminary clarifications about languages and signs;
- explains what arbitrariness is, and how arbitrary conventional language is;
- explains what systematic compositionality is, and how systematically compositional conventional language is; and
- argues for considering the two properties together, and presents one way the two interact in language: the systematic re-use of arbitrary elements.

It then introduces the specific question that this thesis aims to answer – how the systematic re-use of arbitrary elements emerges in a language – and lays out the structure of the remainder of the thesis.

1.1 Language as a set of signs

Human language can be thought of as a set of signs, or mappings between signals and meanings.[1] The word “sign” has been used differently[2], so we provide a few examples of our use here.

1.1.1 Complexity

Words, phrases, sentences, etc. vary in complexity, but are all strings of sounds that map to meanings and therefore are considered signs in this thesis. For example, spoken English includes these mappings:

> "dog" ⇔ a domesticated canine, bred in many varieties
>
> "love" ⇔ to have a deep, tender feeling of affection toward a person
>
> "big house" ⇔ a building in which people live, which is large in size compared to others
>
> "eat pizza" ⇔ to chew and swallow a portion of an Italian open pie made of thin bread dough spread with tomato sauce and cheese

---

[1] The capacity to use language involves more than simply encoding and decoding signs. The meaning of a signal is sometimes just a clue to the speaker’s meaning.

[2] For example, people say there is a “sign for” a certain item. We would say there is a signal for a certain meaning, and reserve “sign” for the pairing of a signal with a meaning.

1.1.2 Modality

People can communicate through various media. Thus, all kinds of signals can carry meaning. In this thesis, I'll provide examples of communication in these modalities:

- Speech. Speech is sound, which fades quickly over time. One can manipulate the pitch, loudness, etc. of it.
- Graphics. Graphics are two-dimensional spatial representations, e.g. drawings and paintings. They have no sound. They do not change over time, except during production.
- Gestures take place in three-dimensional space, and extend into time. They have no sound.

This thesis addresses how the signals of a language relate to their meanings.

1.2 Languages exhibit arbitrariness.

1.2.1 Arbitrariness vs. motivatedness

Saussure (1916) argued that the basic units of language (words) are arbitrary signs - signs in which the connection between the signal and the meaning is arbitrary. In support of this, he noted that different languages have different signals for the same meanings. For example, a horse is called "horse" in English, "cheval" in French, and "caballo" in Spanish[3]. None of these sequences of sounds has an intrinsic relationship with the concept of a horse; there is no natural reason why a particular

[3] The presence of multiple signals mapping to the same meaning isn't a strong argument for arbitrariness. As we will see, there are many different ways (and different degrees to which) a signal can non-arbitrarily relate to its meaning.

signal should be connected to a particular meaning. In fact, any other sequence of sounds would have worked equally well[4].

Arbitrariness is so prevalent in language that one might ask how else a (linguistic) sign could be. In addition to the arbitrary sign (the *symbol*), Peirce (1955) noted two other types of sign: the *index* and the *icon*.

In an index, the signal is directly connected to the meaning in some way (e.g., physically or causally). For example, the presence of smoke means that there is a fire because the fire caused the smoke. Also, someone's crying means that they are in great distress because the distress caused the crying. If the meaning exists, the signal exists – whether or not someone interprets the sign. Thus, pure indices aren't communicative (where communication is volitional); the signals aren't generated for the purpose of communication.

In an icon, the signal resembles the meaning. There are two types of iconicity: imagic (where a single signal physically resembles its referent) and diagrammatic iconicity (where the arrangement of signals resembles the relationship between their referents). A linguistic example of imagic iconcity is onomatopoeia, where speech sounds mimic the sounds of what's referred to, e.g., "meow" and "sizzle". An example of diagrammatic iconicity is Caesar's "veni, vidi, vici", where the temporal sequence of words reflected the temporal sequence of the events they referred to.

Note that this distinction between imagic iconicity (similarity between a signal and its meaning) and diagrammatic iconicity (similarity between the arrangement of signals and the relationship between their meanings) depends on what a basic meaningful signal is. It could be argued that "meow" also illustrates diagrammatic iconicity because the arrangement of "me" before "ow" reflects the temporal order of the sounds an actual cat makes. Perhaps "meow" and "sizzle" both tend to be

---

[4] Of course, a speaker can't utter anything he wants for a given meaning. The sign must be accepted by the other speakers of the speech community. This isn't usually done by some explicit agreement but rather by following tradition – roughly, the sounds you produce when you mean DOG are the same your parents produced.

considered onomatopoeic because at least part of each string of sounds has a natural connection to its meaning *not* in virtue of the arrangement of the sounds. For example, perhaps the “z” sound (or the fact that it is fricative) relates to sizzling. It’s also possible that “meow” is considered an example of imagic iconicity rather than diagrammatic iconicity because it’s mistaken as an irreducible morpheme, i.e. “meow” isn’t obviously composed of meaningful units.

There are many different ways for a signal to have a natural connection to its meaning. Many involve both iconicity and indexicality. Peirce (1955) himself, when introducing the distinctions between icons, indices, and symbols, noted that the categories aren’t mutually exclusive; a sign can be part iconic, part indexical, and/or part symbolic.

To begin with, what is depicted by the signal need not be the actual meaning. If someone wants to refer to Robert DeNiro by drawing, they might draw not Robert DeNiro, but rather a taxi because Robert DeNiro starred in *Taxi Driver*. (Garrod, Fay, Lee, Oberlander, & MacLeod, 2007) To communicate an abstract concept such as love, someone might draw a heart. To communicate the category of vehicle, someone might depict one particular car (an example of the category).

Further, there are often different *ways* to depict any given item. If someone is trying to depict a bowling ball using gesture, they could show the shape of the ball with their hands (perhaps by making an “O” with their two hands). They could also show the motion of the bowling ball (perhaps by tracing a straight, long path on the ground with their finger). They could instead depict the effect the bowling ball has on their body (perhaps by showing the shape their fingers make when they’re in the bowling ball) or the motion they make when they are bowling.

In short, any association one makes with the meaning could be depicted, and this associated item could be represented in a myriad of ways.[5] Since there are many

---

[5] This can be demonstrated. For example, we could task people to communicate a variety of items by drawing, and then show that the drawings for similar meanings (say, the drawings for various people) are not necessarily themselves similar.

different ways for a signal to have some inherent connection to its meaning, where the distinction doesn't matter, I will follow Saussure (1916) and conflate these into the category of "motivated".

Motivatedness (and, thus, arbitrariness) is a matter of degree because the strength of the (inherent) connection between a signal and a meaning can vary across signs. While a sign is often deemed iconic if the signal resembles the meaning at all, iconicity is actually a scale: the more the signal resembles the meaning, the more iconic the sign. For example, a detailed illustration of a professor is more iconic than a sketch of a head with a mortarboard on it. In the context of communication, indexicality is a scale in the same way. For example, if fire always causes smoke and nothing other than fire causes smoke, then there's a very strong connection between smoke and fire.

1.2.2 Measuring arbitrariness

To illuminate the concept of motivatedness vs. arbitrariness, we consider here how to measure how arbitrary a sign is.

Keller (1998) distinguishes between two views of linguistic meaning. Under the representational view, the meaning of the signal is that which it represents, or "stands for". In contrast, with an instrumental notion of signs, one asks how signs function. For example, Keller takes the latter view and characterizes the kind of inference made when interpreting a given signal (causal, associative, or rule-based), rather than characterizing the relationship between the signal and the meaning.[6] This representational/instrumental distinction is helpful as we explore how to measure arbitrariness: we can use linguistic or psychological measures.

---

[6] One advantage to this approach is that it can account for the fact that a sign can function as an icon for one person but a symbol for another. Another, related, advantage is that it can offer an account of how signs can evolve.

Garrod, et al. (2007) use a linguistic measure of the iconicity vs. arbitrariness of drawings. They argue that the complexity of a signal is a measure of the sign's iconicity – all else equal, the less information there is in the signal, the less the signal can resemble its meaning. Note that this measure captures how much information the signal might carry about the meaning, but cannot capture how the item depicted in the drawing relates to the drawing's meaning. Indeed, the meanings of the signal do not factor into this measure at all.

It is possible to construct a measure of arbitrariness that takes account of both signals and meanings. Although iconicity and indexicality are properties of individual signs or constructions, if we found regularities across unrelated languages – certain sounds tending to map to certain meanings – we'd have support that these connections between sound and meaning are universal (and, thus, presumably inherent or natural). Conversely, the fewer the cross-linguistic regularities found, the more arbitrary the connections must be.

Fay, et al. (2008) offer a psychological measure of the relative iconicity/arbitrariness of signs: transparency to naïve overseers. The more motivated a sign is, the more transparent the connection between signal and meaning. When a sign is highly motivated, people who don't know the language should nonetheless be able to guess the meaning from the signal. When a sign is very arbitrary, they should not be able to; people need to be taught arbitrary signs. Thus, we can measure arbitrariness as accuracy on identification of the meaning from the signal by a naïve overseer.

1.2.3 How arbitrary is language?

We've considered how signals *can* relate to meanings. We turn now to how arbitrary or motivated the signs in language actually are. We will see that, while there is some motivatedness in conventional languages, they make use of vast numbers of arbitrary signs.

Several types of diagrammatic iconicity (in which the arrangement of signals is similar to the relationship between their meanings) are exhibited in language and literature. For example, there is Iconicity of Complexity (Haiman, 1985), in which the complexity of a signal – how many signals it's composed of, whether there are any embeddings in it, etc. – indicates the complexity of the concept it refers to. Reduplicative plurals (e.g. "p." as the abbreviation for "page" but "pp." for "pages") are an example: producing the singular form multiple times indicates that there are multiple referents. Also, relativization, subordination, and the like involve an increase in complexity in both signal and meaning.

There is also a kind of secondary diagrammatic iconicity. As in primary diagrammatic iconicity, the relationship between signals is similar to the relationship between their meanings. But while primary diagrammatic iconicity is independent of any particular language (i.e. those relationships are possible in any language), secondary diagrammatic iconicity is not. For example, in Caesar's "veni, vidi, vici", each verb consists of two syllables and each syllable is the same length, formed by the same consonant ("v") and a vowel. Caesar chose these three similar words to emphasize the similarity (in terms of ease of execution) of the three actions they refer to, i.e. that conquering was as easy as coming and seeing for him. But it is only by chance that these three words are similar; it would not work in English. This and other kinds of diagrammatic iconicity are more creative and less conventional. That is, they don't involve rules of the language but rather manipulations of it.

There is some evidence for sound symbolism, i.e. that the individual sounds in words (phonemes) carry meaning. If the sound has a natural connection to its meaning, then the sound symbolism is an example of iconicity. Onomatopoeia, when speech imitates the sound of the referent (illustrated above) is one type of sound symbolism. Many languages have onomatopoeic words referring to anything from animal sounds ("cock-a-doodle-doo") to the noise made in an explosion ("kaboom"). (www.writtensound.com) However, the vast majority of words in any language are not onomatopoeic. We'll discuss sound symbolism further in the next chapter. For now, we note that it's unclear how prevalent such natural connections between signals and meanings are. What is clear is that each of the world's languages has at least thousands of morphemes. Since a morpheme is the basic meaningful unit of a language (a word or word unit), it itself has no meaningful parts (parts from which to derive its meaning), and so it must be learned. (If a morpheme were iconic, it would have smaller meaningful units, and cease to be a morpheme.[7])

Klima and Bellugi (1979) argue that, even in American Sign Language, where one might expect a high degree of iconicity, many of the words are arbitrary - there is residual iconicity but plays no role in the structural descriptions of the language or the way the language is processed.

1.3 Languages exhibit systematic compositionality.

So far we have distinguished between motivated connections between signals and meanings and arbitrary ones; a signal was either constrained by its meaning in virtue of naturalness, or not constrained by its meaning at all (except by convention).

But there is another possibility: a sign can be non-arbitrary without having any natural connection to its meaning, but rather in virtue of other signs in the language. Consider the word "walked." It is not the case than any other signal for that meaning would have worked equally well. Specifically, the "-ed" suffix was determined from the fact that the meaning is past tense. It's not that "-ed" has some natural connection to the past tense - the past doesn't sound like "ed". Rather, the word includes that suffix because other words for past tense verbs do. Here we introduce the concept of systematic compositionality, a type of non-arbitrariness that applies to *sets* of signs.[8]

---

[7] In theory, a morpheme could be motivated in virtue of its being indexical. But pure indices seem to be incompatible with the volitional communication, which this thesis addresses.

[8] Diagrammatic iconicity requires the existence of multiple signs, since the similarity is between the arrangement of signals and the relationship between their meanings. But it is not a property of a set of signs but rather of a single complex sign, akin to a linguistic construction. This is evident from the fact that, the very first time someone produces a diagrammatic icon, it should be understood – the

1.3.1 Systematic compositionality vs. arbitrariness

An expression is *compositional* if its meaning is a function of the meanings of its parts and the way in which they're combined. This principle is commonly attributed to Frege. (Garcia-Alvarez, 2005) For example, the meaning of the utterance "Man bites dog." includes the meanings of "man", "bites", and "dog" – the situation that the utterance refers to involves a man, biting, and a dog. Further, the order of the words indicates that it's the man biting the dog – not vice versa, not someone else biting both of them, not someone biting someone else and the man and dog simply existing, etc.

Compositionality is closely related to *combinatoriality*, in which signal (or meaning) elements recur across signals (or meanings) but do not carry meaning.

How does one know the meaning of the parts of an expression? This is where *systematicity* comes in. "Systematicity" is a broad term, usually used to refer to some kind of regularities across instances. For example, Fodor and Pylyshyn (1988) argue that cognitive capacities are systematic because some are intrinsically connected to each other. They cite an example of a linguistic capacity: across language learners, if one knows how to say that John loves the girl, then he also knows how to say that the girl loves John.

This thesis addresses the nature of the mappings between signals and meanings in communication systems. Thus, we use "systematicity" to mean regularity in the mappings between signals and meanings across signs.[9]

A language can be systematic without being compositional. There can be a common frame of reference across signs that is orthogonal to compositionality. For example,

---

Hearer understands because the connection is natural, not because he has seen the connection elsewhere.

[9] Note that we are not taking account of the tokens of the sign types. We assume that one always means the same thing when he says "red house" and that he always says "red house" when he means that.

imagine that someone has to refer to musical notes, but can only draw on a sketchpad to do so. They might draw a tall line to refer to the highest note, a short line to refer to the lowest, and so on. Even though each sign is simple (i.e. no sign is composed of independently-meaningful parts), there is a kind of regularity across the signs, and the signal for a given meaning is constrained by the existence of other signs.

It's harder to imagine, but the converse is also true: a language can have compositional expressions without being systematic. For example, if one is explicitly taught that "zeit" means time and "geist" means spirit, he can figure out that "zeitgeist" means spirit of the time. This holds regardless of whether he has seen "zeit" or "geist" in any other contexts.

Despite the fact that compositionality and systematicity are independent concepts, it's their combination we're interested in here. Thus, we focus on *systematic compositionality*: the regularity of signal elements and meaning elements across (complex) signs. We consider the extent to which signals for similar meanings share something, and to which the meanings of signals that share something are themselves similar.

Why focus on this composite property – systematic compositionality – rather than on either compositionality or systematicity? While not always acknowledging it, the literature often addresses the combination of the two. For example, both systematicity and compositionality are required for people to be able to generate novel complex signs (e.g., sentences) that can be understood by others. Likewise, some claims about how compositionality evolves depend on the parts of signs recurring, not just on the signs having parts. For example, Kirby (2000) demonstrates that a compositional language can be transmitted to a new generation of learners even though they do not observe every sign in it. Of course, this only holds if the language is also systematic: if someone hasn't learned how to express the specific meaning hat a man bit a dog, they cannot know how to express it unless the way they should express man, biting, and dog in the new context is the same as they observed in other contexts.

How else might sets of signs be? First, sets of signs can have no systematic compositionality at all. In this case, each sign would be arbitrary and thus would map holistically to its meaning. That is, we would have something like a word for every possible meaning we'll want to express, even the meanings composed of meanings that already have signals. For example, we'd have a word to express the meaning that a man is biting a dog, and that word would sound nothing like "man", "bites", or "dog."

Second, a language can exhibit various degrees of systematic compositionality. For example, idioms make a language less systematically compositional. Consider "bought the farm". This is composed of the three parts "bought", "the", and "farm" arranged in a certain order. The phrase should mean that there exists a farm and that someone purchased it. Instead, it means that someone died. Although both the signal and its meaning are complex, the signal maps as a whole onto its meaning – the meaning is not derived from its parts or the way its parts are put together. Note that idioms can't exist without the existence of some degree of systematicity in the language. An idiom is distinguished as meaning the "wrong" thing – it's composed of independently meaningful parts, and one can derive a meaning of the whole expression from them, but this meaning is the wrong one. In other words, "cat", "pat", and "hat" aren't idioms – although each is clearly composed of parts, and each must be learned holistically, the sounds in them never carry meaning.

Irregulars also make a language less systematic, but in a different way. Most past tense words contain "-ed" but "went" does not. Irregulars don't contain independently meaningful parts, but should.

Having unnecessary words might be a special case of irregularity. It's logically possible that a language has morphemes only for basic semantic features. Anything that could be uniquely described by listing its semantic features would be. For example, if there's only one yellow fruit in the world, there's no need for the word "banana" if the words "yellow" and "fruit" already exist (and occur in other complex signs, contributing consistent meanings).

1.3.2 Measuring systematic compositionality

As with motivatedness vs. arbitrariness, we explain systematic compositionality further by considering how it could be measured.

Kirby (2000), presenting simulations of the evolution of systematic compositionality, compares the number of rules in an agent's grammar to the number of meanings that that agent can express. An agent's language is systematically compositional to the extent that the agent stores fewer rules than the meanings that he can express – his language cannot be a list of unanalyzed vocabulary items for complex meanings.

Brighton, et al. (2005) define compositional languages as topographic mappings from signals to meanings. That is, compositionality is defined as the degree to which similar signals map to similar meanings. "Man bites dog" and "man bites cat" are similar signals because they share something, viz. "man bites". Because these signs are compositional[10], their meanings are also similar – the meaning of each sign is derived from the meaning of its parts and the way they're put together so, since they share parts in the same position, their meanings are similar. In contrast, "kicked the bucket" and "kicked the pail" are similar forms, but their meanings are not similar. The similarity of the two forms doesn't indicate similarity between their meanings. Thus, how topographic the set of mappings between signals and meanings is an indication of how compositional the language is. Compositionality can then be measured as Pairwise Distance Correlation, the correlation between the distance between pairs of meanings and the distance between the corresponding pairs of signals. (Brighton, et al., 2005) In topographic mappings, there will be a positive correlation.

RegMap (Tamariz & Smith, 2008) is a measure of the regularity in the mappings from signals to meanings (or vice versa). For each possible mapping from signal

---

[10] To be clear, compositional languages must also be systematic in order to be topographic mappings.

component to meaning component, one compares the number of examples of that mapping with the number of counterexamples to it.

Both Pairwise Distance Correlation and RegMap are linguistic measures, characterizing how signals relate to meanings. We might instead take a more psychological approach. For example, given that compositionality is defined in terms of how the meaning of a complex sign is derived, we could probe how people actually process complex signs. This way, we could distinguish between regularities that exist in a language and regularities that people actually use.

In addition to signal comprehension, signal production could be used as a measure of systematic compositionality. If the systematic compositionality has a psychological reality, when someone generalizes to a new meaning, their signal will include the elements they observed in signals for similar meanings.

1.3.3 How systematically compositional is language?

Language does have idioms and irregularity. Also, there appears to be much in language that could be systematically compositional but isn't; it's not the case that only what cannot be described compositionally is expressed with an arbitrary sign. Nonetheless, the degree of systematic compositionality in language is striking, especially when we consider the infinite set of grammatical sentences we can produce and understand.

1.4 Considering arbitrariness and systematic compositionality together

1.4.1 Why look at the two together?

We began by defining motivatedness and arbitrariness as properties of individual signs – roughly, whether there's any kind of natural/inherent connection between a signal and a meaning. We then introduced systematic compositionality as a property of sets of signs – roughly, the degree to which signal elements map to the same meaning elements across complex signs. If one dimension is a property of individual signs and the other is a property of sets of signs, it's tempting to view them as orthogonal, and dismiss the relationship between the two is trivial.

However, although iconicity and indexicality are properties of individual signs, it's sensible to talk about the relative motivatedness or arbitrariness of a set of signs. It's simply the sum of those measures on the relevant subset of individual signs. For example, if all the signs in a language are highly iconic, it's sensible to call the language highly iconic. Likewise, although systematic compositionality is a measure on sets of signs, it's sensible to describe an individual sign as systematically compositional. If there is some systematic compositionality in a set, the systematic compositionality of an individual sign is how well it fits into the identified system. For example, the sign (compound word) "greenhouse" is partly systematically compositional and partly arbitrary – it's something like a house and is often green, but it has a more specific meaning than just "green house". In the end, the two dimensions are both characterizing what constrains the signal one produces for a given meaning, and what constrains the meaning of a given signal.

These properties – motivatedness (where the signal has an inherent connection to its meaning), systematic compositionality (where parts of the signal map to parts of the meaning in the way they do in other signs), and arbitrariness (where the signal is not related to the meaning in either of these two ways, but likely works just because of convention) – all apply to the signals for the same meanings. If there were a level of semantic representation (e.g. of features like red, person, etc.) that got motivated or arbitrary signs and then higher levels of semantic representation (e.g. things that can be described as sets of these features, like firefighter) that were systematic, we could address the two dimensions independently. But this isn't the case. Signals for the same level of representation can be iconic, systematically compositional, or completely arbitrary.

By looking at the two dimensions in terms of constraints on individual signs (rather than as properties of individual signs vs. properties of sets of signs), we can address *why* a sign takes the form that it does. Consider briefly a set of iconic drawings for concepts related to primary education: teacher, classroom, teaching, and school. Since the drawer intended the drawing of each concept to resemble that concept, it's plausible that each would contain a drawing of a blackboard. The set of drawings would be technically systematic then. Yet, one might be resistant to calling them systematic, because that label obscures what constrained what was drawn. In other words, in this case the drawer didn't draw a blackboard in any one drawing in order to be consistent with this other drawings; the systematicity was coincidental. Now consider that this same person produced a drawing for school bus as well, and it included a chalkboard. Now we might argue that it is systematicity, not iconicity, that's constraining these drawings. By considering the motivated vs. arbitrary and the systematically compositional vs. arbitrary dimensions together, we are able to understand what's constraining the production of each individual signal.

1.4.2 How the two dimensions generally interrelate

Now that we've established that motivatedness, systematic compositionality, and arbitrariness each characterize the constraints on the signal one produces for a given meaning (and on what meaning one gets from a given signal), let's take a closer look at how they interrelate. In what follows, we'll assume a certain level of semantic representation, and look at what constrains the signal one produces for a meaning at that level.

In some ways, motivatedness is compatible with systematic compositionality. First, one could argue that iconicity is a type of compositionality. Iconicity is defined in terms of the *way* a signal relates to its meaning: does it resemble it or not? In this sense, a drawing of a professor that consists only of a drawing of a mortarboard is iconic. But in order to resemble complex meanings, iconic sign will tend to consist of independently meaningful parts, i.e. be compositional. Further, a set of *maximally* iconic signs will be systematic. To the extent that there are more iconic ways to represent given semantic features, then the same features will be represented the same way across signs. For example, the most iconic way to represent classroom might be to draw what the inside of a typical classroom looks like, and the most iconic way to represent teacher is to draw what a typical teacher looks like (including the environment in which a teacher is typically found: a classroom), so the drawings for these two related meanings will share many elements.

In other ways, motivatedness and systematic compositionality can be seen as opposing pressures. For example, systematicity is not necessarily iconic, because it is less constrained semantically: an element that recurs across signs can mean anything; it need not map to perceptual features of the referents, but could mean something more functional or abstract. If someone is drawing an airplane, iconicity and systematicity might pull them in different directions – they can either produce a drawing that looks like an airplane or produce one that looks more like the other vehicles that have been drawn.

1.4.3 Languages systematically reuse arbitrary elements.

There could be relatively arbitrary or relatively systematically compositional signs at all but the most basic levels of linguistic analysis, including words, phrases, etc. We consider here just one case of their interaction: the systematic re-use of arbitrary elements. For example, the words “big”, “red”, “apple”, and “house” are arbitrary, and the referring expressions “big house”, “red house”, “big apple”, and “red apple” systematically re-use them.

This property – the systematic re-use of arbitrary elements – is closely related to Duality of Patterning. (Hockett, 1960) Duality of Patterning refers to the fact that, in human language, meaningless units combine to create meaningful units that also combine. Specifically, morphemes are combinatorial - individual sounds recur across morphemes but do not carry any meaning - and then multi-morpheme words or phrases are compositional – morphemes recur across words and phrases and do carry the same meanings when they do. The systematic re-use of arbitrary elements is more general – we address how meaningful, arbitrary units combine, but do not address whether those units are combinatorial.

1.5 How can the systematic re-use of arbitrary elements evolve in language?

Interestingly, the two features we’ve been talking about – arbitrariness and systematic compositionality - are often cited as hallmarks of human language. For example, Hurford (2004) notes that “Human language is qualitatively different from animal communication systems in at least two separate ways. Human languages contain tens of thousands of arbitrary learned symbols (mainly words)... Human language also has complex compositional syntax. The meanings of our sentences are composed from the meanings of the constituent parts (e.g. the words). “ Given the uniqueness of the systematic re-use of arbitrary elements, the rest of this thesis addresses how it can arise in a human language.

To answer this, we will survey the literature on the question in Chapter 2. In Chapter 3, we point out some issues about the general approach that previous work has taken, and offer an alternative. Chapters 4, 5, and 6 present experiments that probe the systematic re-use of arbitrary elements in novel sign systems. Chapter 7 concludes by summarizing the main contributions of the current research, and making suggestions for future work.

Chapter 2

Previous work on how the systematic re-use of arbitrary elements can arise in a language

In Chapter 1, we laid out the key theoretical concepts surrounding the systematic re-use of arbitrary elements and motivated the specific question of how it can arise in a language. Here we survey the literature on this question.

2.1 What we know about novel signs

2.1.1 People can find ways to successfully communicate in the absence of conventional communication systems.

Sometimes, people are not exposed to any language. In particular, deaf individuals have often found themselves in situations in which they needed to create a language from scratch. Despite this, people find ways to successfully communicate. For example, when a deaf child is born to hearing parents who don't know sign language, he isn't exposed to any conventional language. Goldin-Meadow, et al. (1995)

showed that, despite the lack of language input, these children create a system of gestures to communicate with their caretakers, called homesign. As another example, the Al-Sayyid Bedouin group is an isolated community with an unusually high proportion of deaf individuals. Sandler, et al. (2005) describe the Al-Sayyid Bedouin Sign Language (ABSL) that has arisen from scratch in this community.

Recent experimental work has pioneered the controlled study of how communication systems emerge and evolve. One goal of this work is to explore how people find ways to successfully communicate in the absence of conventional communication systems, especially by creating novel signs. For example, de Ruiter, et al. (2007) had subjects play a video game in which they had to cooperate in order to move their tokens into the correct orientation and location. Similarly, partners in Scott-Phillips, et al. (2009)'s game were to move to rooms of the same colour, but could not see the colours of each other's environments. In both cases, although players had only the motion of their own tokens with which to communicate, most were eventually successful.

2.1.2 The first signs people produce are motivated.

When people find themselves in the absence of language, the first signs they produce are usually iconic or motivated in some other way, i.e. *not* arbitrary[11]. This is very clear when people invent signs in the manual modality, i.e. gestures. For example, in homesign, the shape and motion of an object are encoded in the shape and motion of one's hands. (Goldin-Meadow, et al., 1995)

We see this same thing – that the first signs people produce are motivated – in the experimental work. For example, the task of Galantucci (2005)'s subjects was to bring their agent into the same room as their partner. Each of the rooms was marked with an icon, but partners could communicate using only a novel graphical channel that transformed their signals in an obscure way. After a number of attempts, some subjects established signs that allowed them to win the games. The signals for the

[11] Scott-Phillips, et al. (2007) offer an interesting exception.

rooms varied widely across the games, but many had inherent connections to their meanings: representing the room's location in the environment or the shape or the number of vertices of the room's icon. People can find motivated ways to graphically represent even music. Healey, et al. (2007) presented a Music Drawing Task, in which subjects refer to short pieces of unfamiliar music with their partners. They can draw on a virtual whiteboard, but may not write letters or numbers. Most of the signals they produce to solve the task make use of one of two strategies. Some drawings consist of recognizable objects, figures, or scenes – these are iconic drawings of associations one might make to the pieces of music, e.g. a race car for a piece of music with a fast tempo. Other drawings represent the structure of the pieces, e.g. drawing a curve with peaks and valleys that match the high and low notes in the piece. Clearly, people can exploit a wide range of connections between signals and meanings to get communication systems off the ground.

While we see that novel signs are motivated when people invent gestures and in the experiments using novel communication channels (often graphics), the reader might wonder how the beginnings of spoken language could be motivated. First, as we saw in Chapter 1, sometimes a word can sound like what it means. For example, many languages have onomatopoeia. It's easy to imagine how, even before any language has been established, someone could refer to a bear by producing a "grrr" sound.

But it's justifiably hard to imagine being able to sound like *everything* one might refer to,[12] as some things one will want to refer to may not have a distinctive sound. However, sound can carry meaning in many other ways. To start, many languages exhibit sound symbolism at linguistic levels below the word (e.g., clusters of phonemes, or even phonetic features). For example, Parault and Schwanenflugel (2006) confirmed the psychological reality of seven sound symbols in English:

- "fl-" meaning moving light or movement in air,
- "gl-" meaning unmoving light,
- "j-" meaning up and down movement,
- "scr-" meaning grating impact or sound,

[12] We'll see later that this isn't necessary in order to get language off the ground.

- “sl-“ meaning smoothly wet,
- “sn-“ meaning breath-noise or quick movement or creep, and
- “sw-“ meaning swift movement.

Participants were asked to generate a written definition for obsolete English words that either used one of the sound symbols or were non-sound symbolic. Then, for each word, the participants had to choose the correct definition from among three distractor definitions. Sound symbolic words yielded significantly more accurate definitions than non-sound symbolic words, and the correct definition was recognized significantly more often for sound symbolic words than for non-sound symbolic words. Thus, these sounds carry meaning.

In a similar vein, Monaghan and Christiansen (2006b) demonstrate that there is a systematic relationship between phonological features of words and their lexical category. They looked at the most frequent nouns and verbs in English and French, and found that in both languages the first few phonemes and especially the last few phonemes correctly classified the words as noun or verb significantly above chance. Monaghan and Christiansen (2006a) extend this work to other languages and another aspect of the meaning of words. They analyzed frequent words in English, Dutch, French, and Japanese and found that a number of phonological cues in each distinguished function words from content words and nouns from verbs.

Syntactic patterns can also carry meaning, independent of the words used in them. (Goldberg, 1995) For example, it’s implausible to claim that the meaning of sneeze includes causing motion and yet it appears to in a sentence like “He sneezed his tooth right across town.” (Goldberg, 2003) It’s the number of arguments that “sneezed” appears with that gives it this meaning. The syntactic patterns in which a verb appears provide information about its meaning, since its meaning must be consistent with the patterns’ meanings. Fisher, Gleitman, and Gleitman (1991) demonstrate the existence of a number mappings between the syntax and the semantics of verbs - verbs that were judged to share certain syntactic properties were also judged similar in meaning. For example, mental verbs such as “believe” are associated with the sentential complement pattern (e.g., “She believes that the apple is rotten”). Lidz, Gleitman, and Gleitman (2003) suggest that some of these mappings between syntactic patterns and the meanings of words may even be universal. They show that, across divergent languages (Kannada and English), child native speakers used the number of arguments of the verb to derive an aspect of its meaning, its causitivity.

Similarly, such sound-meaning correspondences may be universally understood. For example, Japanese has words that mimic voices and also words that mimic manners or states (which don't necessarily have sound). Iwasaki, et al. (2007) showed that native English speakers, with no previous experience with Japanese, are sensitive to aspects of the meaning of some of these mimetic words. They had native English speakers and native Japanese speakers rate various Japanese mimetic words for laughing and walking on several semantic dimensions each. For example, each participant had to rate "yota-yota" (the manner of walking with heavy, faltering steps) on semantic dimensions such as from purposeful to aimless, from big strides to small strides, and from graceful to clumsy. Surprisingly English speakers' ratings agreed with Japanese speakers' ratings for many of these semantic dimensions.

Imai, et al. (2008) further explored the non-arbitrary mappings between sounds and manners of action. They created *novel* Japanese mimetics expressing different manners of walking, based on previous work describing what meanings the individual sounds carry. For example, they created the novel mimetic "hyaihyai" for semi-swift walking with light, playful steps, as ''h" has been shown to express weakness and unreliability and ''y" has been shown to express leisurely, unreliable motion. For each novel mimetic, they created two video clips with a character walking in a manner that either matched or did not match the mimetic. English speakers with no prior experience with Japanese then rated how well the mimetic matched the action in each video. The English speakers rated the matching actions as a significantly better fit than the non-matching actions. Other English speakers heard each novel mimetic and were asked to choose, between the matching and non-matching videos for it, which action the word referred to. They chose the matching action at a level significantly above chance. Thus, even people who don't know any Japanese have some information about how the sounds map to manner in it. This suggests that the mimetics found in Japanese use associations between sounds and meanings that may be universal.

One can also find sound-meaning correspondences that are actually instantiated cross-linguistically. For example, Nuckolls (1999) reviews cross-linguistic evidence that high front vowels (such as in “heed”) tend to express diminutive concepts such as physical smallness, quickness, brightness or lightness, singularity, and attitudes such as affection.

By surveying the world’s languages, we find surprising ways that sounds – from whole words down to phonetic features – carry meaning. We can also probe the associations people make to sound experimentally.

First we note synaethesia, a neurological condition in which people make odd cross-modal associations, e.g., see certain words in certain colours. Ramachandran and Hubbard (2001) suggest that synaethesia may be the extreme form of normal cross-modal associations which helped to get spoken language off the ground.

Marks (1974) began to probe this by presenting (normal) participants with squares of gray paper that varied in their brightness (from black through white). For each, the participants varied either the pitch or the loudness of a tone until it “matched” the brightness of the gray. Participants reliably associated brighter values with higher pitched and louder tones.

Marks (1987) used a classification task to reveal associations between sound and vision. He created stimuli that had both an auditory and a visual component. Each participant’s task was to make a binary judgment about one of them (i.e. to ignore the other). For example, the participant might see a light while hearing a tone, and have to judge whether the light was dim or bright, but make no judgment about the tone. They found that lightness and brightness affected the response time to pitch and, conversely, pitch affected the response time to lightness and brightness. For example, participants classified the lightest stimulus (white) significantly faster when it appeared with a high-frequency tone than with a low-frequency tone, indicating that light visual stimuli are congruent with high-frequency tones. They found other cross-modal interactions: between the loudness of a sound and the brightness of an image, and the pitch of a tone and the shape (round vs. sharp) of an image.

Using a similar paradigm, Ben-Artzi and Marks (1999) explored the non-arbitrary connections between the (spoken) words "high" and "low", pitch, and position. They created eight three-dimension stimuli: the words "low" or "high" spoken in either a low or high pitch presented to the participant through either low or high speakers. Participants were instructed to classify the stimuli as high or low according to just one of the dimensions. In general, participants were quicker to classify when the values on the irrelevant dimensions were congruent with the value on the relevant dimension. For example, classification by position was significantly faster when the pitch and the word were congruent with it (viz. also high or low) than when they were incongruent.

Reilly, et al. (2008) demonstrated that there are non-arbitrary connections between colours and tones. They paired colour swatches that varied in hue and brightness with tones (which vary in frequency). Participants answered whether or not the two were a good match. Participants matched hue to the frequency of the tone linearly: the higher the wavelength of the colour, the lower the frequency of the tone. For example, the vast majority of participants endorsed the match of a 190 Hz tone with a dark blue colour swatch. There was also an interaction between brightness and frequency of the tone. Reilly, et al. also paired novel objects with pseudowords and found interactions between the phonology of the pseudowords and the hue and animacy of the objects.

Simner, et al. (under review) demonstrated non-arbitrary relationships between features of speech and types and concentrations of taste. Participants received drops of 12 tastes: sweet, sour, bitter, and salty at three different concentrations each. Their task was to adjust settings on four different sound continua to match each taste. The four sound continua represented qualities of speech: vowel height, vowel backness, voice discontinuity, and overall acoustic energy. They discovered a preference for mapping increasing concentrations of taste to certain phonetic features: lower, more front, and more staccato vowel sounds, as well as sounds with higher overall acoustic energy. Participants systematically mapped *types* of taste to certain phonetic features as well. For example, sweet tastes were judged to have higher vowels than bitter tastes, to match more continuous vowel sounds than salty tastes, and to have lower overall acoustic energy than bitter, salty, and sour tastes.

Thus, there are many ways vocalizations can be motivated. So far, previous work has illuminated non-arbitrary relationships between sound and at least the hue, brightness, and lightness of colour; spatial position; animacy; diminutiveness (smallness, quickness, singularity, affection, etc.); shape; the type and concentration of taste; and various manners of activities (laughing, walking) and movement (e.g., of light).

If vocalizations still seem to be limited in their expressiveness, note that we need not assume that the beginnings of spoken language consisted of only spoken signals. As Bickerton (2003) notes, language probably began as "a mixture of anything that might serve to convey meaning. The original mixture of isolated grunts and gestures may have eventually settled on the vocal mode merely through the exigencies of communicating at night, over distance, or in dense vegetation." (p. 81) If this is the case, people could use more than speech to convey their meanings. Even if it seems like speech could not be very motivated, people could add information using other modalities to establish these speech signs.

In order to show that the first signs people create can be motivated, we have shown that people exploit inherent connections between the meanings they want to express and the signals they can produce. It's important to note that one need not signal the actual meaning. There may be many *associations to* a given meaning that one can encode in order to convey that meaning. For example, in a game of Pictionary, one might draw a taxi for Robert DeNiro because Robert DeNiro starred in the movie *Taxi Driver*. (Garrod, et al., 2007) That's not a drawing *of* Robert DeNiro, but rather of an association to him. Examples like this can be found in homesign as well. Homesign gestures don't always represent the perceptual features of the object, but sometimes of the hand as it grasps the object or the path of motion of the object instead. (Goldin-Meadow, et al., 1995) Also, most of the homesign systems in

Goldin-Meadow, et al. (1995) had at least one category of event that wasn't based on perceptual features, but rather on a semantic factor (such as vehicles or animate objects). Thus, even though someone might not be able to *sound like* (or look like, or act like, etc.) a particular meaning he'd want to express, he could sound like something associated to it.

2.2 What we know about how signs can change

If the first signs people create are motivated, how do they become arbitrary? We turn now to what we know about how signs change.

Various kinds of semantic change have been documented, in particular by Bloomfield (1933). As an example of just one type of change, the meaning of the word "meat" narrowed – it used to mean food in general but now includes only edible flesh. Different forces for change have also been identified. For example, Grzega and Schoener (2007) name, among many other factors, the increasing dominance of a prototype. One example is how "kleenex", the name of a particular brand of tissue, became the prototypical example of tissues and is now used to refer to any kind of tissue. An overarching theme in the literature on semantic change is that people extend their use of a particular signal to meanings that are similar to the original in some way, as opposed to extending the signal to unrelated, randomly-selected meanings.

Usage-based models of language (Bybee, 2006; Croft, 2005), cite frequency as a major pressure for language change. In these models, someone's grammatical knowledge does not consist in a static collection of rules, but rather is continually shaped by their experience using language. The frequency of signs plays a leading role in language change: frequently-used signs are entrenched, affecting how they're processed. For example, Bybee and Scheibman (1999) show that the word "don't" is reduced (phonologically) most in the most frequent environments in which it appears, such as in "I don't know."

In addition to observing how conventional languages have changed, we can explore experimentally how people adapt their use of conventional language. A large body of research has demonstrated how interaction in a dialogue constrains the production and comprehension of referring expressions, allowing communication partners to converge on linguistic descriptions of items (by converging on conceptualizations of them) and, sometimes, shorten those descriptions. (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987; Garrod & Doherty, 1994; Pickering & Garrod, 2004; Schober & Clark, 1989)

2.2.1 Spotlight on graphical communication experiments

The Concept Drawing Task (Garrod, et al., 2007; Healey, Garrod, Fay, Lee, & Oberlander, 2002) extended this work by prohibiting subjects from using conventional language. This approach is particularly appropriate when investigating the emergence of a property that conventional language already possesses, such as the systematic re-use of arbitrary elements. In this way, we can see what signs subjects create and how those signs evolve. Each round, one of the subjects (the "Drawer") has an ordered list of concepts and the other (the "Matcher") has an unordered list of the same concepts. The Drawer takes each of the concepts in turn and produces a sketch for it on virtual whiteboard they share, so that the Matcher can identify the concept. Players can draw whatever they like but may not write anything, such as letters or numbers. Pairs play for a number of blocks, so that each concept is drawn and identified several times over the course of the game. We can then examine how the signals for the meanings changed over time.

2.3 What we know about getting arbitrariness

Motivated signs can become arbitrary through interaction. Among other measures of how the signals for the items changed with interaction, Garrod, et al. (2007) found that graphical complexity (roughly, the amount of virtual ink used) decreases. They argue that this complexity is a measure of iconicity – all else equal, the less information there is in the signal, the less the signal can resemble its meaning – thus showing that signs become more arbitrary with interaction.

Fay, et al. (2008) measure the motivatedness vs. arbitrariness of signs as transparency to naïve participants. They find that the signs produced by pairs of participants in a community retain their transparency over the course of interaction better than the signs produced by isolated pairs.

This move to arbitrariness is also seen in sign languages. Frishberg (1975) argues that there was a strong tendency for signs to become more arbitrary rather than maintaining a level of iconicity in American Sign Language (ASL). For example, the sign for "we" was a series of separate thrusts, first at one's own chest and then at three or four other persons (real or imagined) and then at one's own chest again. Today, the sign consists of two touches on the chest, with a small, smooth sweep of the wrist or arm between the touches. This lexicalization represents a move "from a composite, explanatory, and iconic representation of the notion of 'we,' to a conventionalized, more arbitrary form." In fact, Goldin-Meadow, et al. (1995) argue that, for all gestural communication systems, iconicity acts as an organizing principle only in the earliest stages.

2.4 What we know about getting systematic compositionality

2.4.1 There are suggestions of systematic compositionality in novel sign systems, but no measures of it.

We now see how arbitrariness can emerge in a novel communication system; what about systematic compositionality? These above studies don't measure the degree to which signals for similar meanings share an element.[13] Some provide examples of novel sets of signs that appear to exhibit systematicity (De Ruiter, et al., 2007; Galantucci, 2005; Healey, et al., 2007), but the recurring elements appear highly iconic. That is, we don't find the systematic re-use of *arbitrary* elements there.

Garrod, et al. (2007) provide an example of the systematic re-use of arbitrary elements: one pair's successive drawings for "art gallery". Each contains three elements: an iconic drawing of a building, something like a "less than" symbol, and an iconic drawing of a painting. They note that the building element became

---

[13] Healey, et al. (2007) offer a measure of a sense of "systematicity" (viz. the use of a common frame of reference when producing signals) that is more general than ours here.

associated with all and only the institutions in the set of concepts, and describe how the building and painting elements simplified to become more arbitrary. As they don't actually measure systematicity, it's left open how common this systematicity is across pairs, and how early this systematicity emerges with respect to the evolution of arbitrariness. For example, did the building element occur in drawings for institutions other than art gallery (say, for parliament) while it was still very iconic? Or did the pair change its drawings for the other institutions later in the game?

2.4.2 Synthetic and analytic protolanguage theories

Language evolution researchers theorize about what a protolanguage could have looked like, and how it then could have evolved into a proper language. Hurford (2000) identifies two dominant views of the move from protolanguage to language: synthetic and analytic. On the synthetic view, the signals in protolanguage had atomic meanings. These were then strung together to make the kinds of phrases and sentences we find in language. On the analytic view, the meanings of the signals in protolanguage were complex. They were then dissected to produce signals for atomic meanings, more like the words we find in language.

Wray (1998) takes the analytic view. Specifically, she envisages the signs in protolanguage being processed holistically, much as idioms (e.g., "that's the way the cookie crumbles") are today. She suggests that the signals would share phonemes with each other and the meanings of the utterances would share elements. For example, "mebita" might mean GIVE HER THE FOOD, "ikatube" might mean GIVE ME THE FOOD, and "kameti" might mean GIVE HER THE STONE. By chance, "mebita" and "kameti" share a syllable and also share an element of their meaning. A language learner analyzing these utterances could infer that "me" means HER. Generations of learners could chip away at the analysis of the language. Tallerman (2007) argues that this holistic approach is fundamentally flawed. One problem she notes is that, since recurrences between elements of the signal and elements of the meanings are chance, there should be at least as many counterexamples as there are examples of any putative word. Some of these criticisms are addressed by K Smith (2006). Regarding the number of counterexamples outweighing the number of examples, he notes that whether it's a problem depends on how the language learner deals with counterexamples – perhaps people analyzed the utterances whenever they noticed a recurrence, and simply ignored counterexamples to it.

Debates over the nature of protolanguage are heated, but some work tries to bridge the gap. K. Smith (2006) proposes that the analytic and synthetic views may be compatible with each other, if protolanguage consisted in a series of stages. In a similar vein, A. Smith (2008) suggests that protolanguage utterances had varying levels of semantic complexity. As such, people used both concatenation and segmentation.

With respect to systematic compositionality, we note another way the views converge: in both, the mappings between signals and meanings in protolanguage are assumed to be arbitrary; protolanguage is assumed not to be systematically compositional. To foreshadow the approach taken in this thesis, we suggest another possibility: that novel signs are systematically compositional, and they are so because novel signs are motivated and motivatedness often implies concatenation.

2.4.3 Spotlight on the Iterated Learning Model of the evolution of compositionality

The Iterated Learning Model (Brighton, et al., 2005; Kirby, 2000, 2001; Kirby, Cornish, & Smith, 2008) demonstrates how a set of arbitrary signs could become systematically compositional. Here's how it works in a nutshell: even in a set of arbitrary signs, the signals for two similar meanings (e.g., that John loves Mary and that Bill loves pizza) might share something (e.g., the syllable "ka"). A language learner might note this mapping between signal element and meaning element. Now, a person doesn't learn signals for every meaning he'll want to express in his lifetime. (For example, I want to express that I'm writing a dissertation on non-arbitrariness in novel communication systems, but no one ever expressed that exact meaning to me.) When one wants to express a meaning he hasn't learned a signal for, if that meaning contains the same element (here, loves) as the mapping, he could generate a signal that includes the signal element (here, “ka”). Over generations of people replacing arbitrary signs this way, the language becomes more systematically compositional.

The model is simple. There’s a set of possible meanings. In one implementation, each is represented as a predicate-argument proposition taking a limited set of atomic arguments, like eat(tiger, john). In another, each meaning is an item with one of three possible shapes, one of three possible colours, and one of three possible motions.

There’s also a set of possible signals. These are strings of various (but usually limited) lengths composed of letters or syllables from a finite set.

Crucially, the set of meanings and the set of signals are both combinatorial. This makes it possible for two meanings to share an element, for two signals to share an element, and for the language to be systematic compositional (where signals that share an element map to meanings that share an element).

In its simplest implementation, the model starts out with two agents: A and B. (For ease of explication, A is male and B female.) Agent A produces signs; he’s assigned a random proper subset of the possible meanings and, for each, generates a random signal. Agent B learns all of his signs, i.e. she stores all the mappings between signals and meanings he produced. For example, A might produce, and B might learn, a set of signs including:

sdx ⇔ loves(john,mary)
lkq ⇔ loves(mary,john)

Crucially, B generalizes from the signs whenever she can. By chance, two signs might permit a generalization: the signals for two similar meanings, i.e. meanings that share an element, might have also shared an element, as in:

filg ⇔ admires(mary,john)
finv ⇔ admires(pete, john)

B can spot the mapping between the signal element and meaning element that recurs in these two signs, and stores it[14]:

fi X ⇔ admires(X, john)

She then simply creates mappings between what's left of the signal and what's left of the meaning, for each sign:

lg ⇔ mary
nv ⇔ pete

Eventually, A dies, B becomes a sign producer, and a new sign learner (C) is born. Like A was, B is now assigned a random proper subset of the possible meanings to express. But unlike A, B knows some signs already. If B has learned a mapping between the meaning she is to express and a signal, she utters that signal. If she has learned no mapping for any part of the meaning, she utters a random string. She can also use what she learned from generalization when producing a signal for a meaning. If she has learned a mapping between part of that meaning and a (sub-)signal, she utters that and then utters random strings for the rest of the meaning.

The model goes on for many generations. In each, the producer dies, the learner becomes the producer, and a new learner is born.

This Iterated Learning Model work consistently finds that systematic compositionality increases over generations. (We have already reviewed the ways that systematic compositionality is measured in this Iterated Learning Model work, in Chapter 1.) This works because the signal-meaning mappings in a compositional language are more general than those in an irregular language. A mapping between a part of a string and a part of a meaning can be observed in more than one (complex) sign, whereas an irregular mapping between a complex signal and a complex meaning will necessarily be observed less frequently. For example, one is less likely

[14] The actual rules in this particular implementation of the Iterated Learning Model are more complex, taking account of grammatical categories. We simplified them here because we are interested in compositionality only.

to have heard someone express that they are writing a dissertation on non-arbitrariness in novel communication systems than they are to have heard someone express anything at all about a dissertation. This matters when learners observe only a subset of the language, because then the mappings between signal elements and meaning elements are more likely to be observed and used in the next generation. Conversely, irregular rules will die out unless they happen to be observed by every generation.

The Iterated Learning Model can also explain how languages can retain a degree of irregularity. Kirby (2001) introduced noise to the signals produced, so that part of the signal was deleted at random. This would sometimes make a regular signal irregular. In addition, the frequency distribution of the meanings expressed became non-uniform (so that some meanings were expressed more often than others). Irregular signals that were mapped to frequent meanings remained in the language because they were observed and subsequently used by every generation.

In sum, the Iterated Learning Model starts from a set of signs that are arbitrary but combinatorial: the set of signals is combinatorial, the set of meanings is combinatorial, and mappings between the signals and the meanings is arbitrary. By chance, similar signals will sometimes map to similar meanings. Agents notice these chance recurrences between signal elements and meaning elements, and use them when generating signals for meanings (that they have not learned signals for). An agent will learn the signals for only some of the meanings he will eventually need to express. This provides the opportunity for each generation to replace irregular signs with more compositional ones.

One might doubt how well this model matches the behaviour of real human learners. However, Kirby, et al. (2008) demonstrates that systematic compositionality evolves in generations of human experimental participants as well. Each participant learns an "alien language" and then is tested on it. The language consists of words for a highly structured set of items (27 items, where each is one of three shapes, is one of three colours, and has one of three motions). The words that the first participant learns are also highly structured; they're randomly generated strings of syllables, but there are only 9 syllables used. Each subsequent participant learns the words that the previous participant produced during his test. There is an observation bottleneck: each participant is trained on only half of the language but tested on the whole language.

Kirby, et al. (2008) examine how the languages change as they're learned and reproduced by each participant in the chain. They find that the languages become easier to learn: the language that the last participant produces during her test matches her input language (of which she's observed only half) better than the language the first participant produced did. Further, the languages become increasingly structured: similar meanings are expressed using similar signals to a greater extent in the last generation than in the first generation.

2.4.4 The emergence of systematic compositionality is tied to the need to express new meanings.

Selten and Warglien (2007) confer that compositionality is tied to the need to express new meanings. They ran a communication experiment where the meanings are highly structured: each item is one of a few shapes, with one of a few inserts (or none), with one of a few colours (or not coloured). These semantic features are introduced in stages. For example, the items referred to in the first rounds are just shapes, but in the next round they're shapes with inserts. The possible signals are also highly structured – each round, the participants are allowed to construct signals from just a few letters. There is usually one letter per feature value. For example, if there are three possible shapes, three possible inserts, and three possible colours, there will be nine letters. There's a pressure to be compositional because there's a cost for using each letter, so one is motivated to avoid long signals. There's also a pressure to converge on signs with one's partner because that's how one earns points.

They find that compositionality has some benefits. For example, pairs with compositional grammars in the middle of the game were able to extend them when they had to assign names to new items. Pairs with noncompositional grammars weren't. Also, compositional signs are more successful in environments with novelty. In one condition, there's a round where every item is a brand new combination of old features. The pairs who had compositional signs scored more points. That suggests that compositional grammars balance the pressure to reduce production effort with the pressure to converge on a set of signs with one's communication partner better than ungrammatical communication systems.

They also find that compositional systems are more likely to arise in certain conditions. Specifically, pairs assigning signals to novel items are more likely to develop compositional grammars than those assigning signals to items they had seen before (but not with each other - each had played with a different partner in the previous session).

2.4.5 Homesign

Thus far, we have seen that work on novel signs tends not to address systematic compositionality directly, and work on systematic compositionality starts from artificial – not novel – signs and thus cannot address arbitrariness directly.

Goldin-Meadow, et al. (1995) aim to capture a longer timeframe, addressing how systematicity arises in homesign. They studied the gesture systems of four homesigning children from when they were around three years old to when they were around five years old, over the course of seven sessions. Each gesture had a handshape and motion component, and they found that most gestures (i.e. combinations of specific shapes and motion) in the beginning of the study didn't refer to more than one event. By the end of the study, however, the children did extend gestures to more than one event. Specifically, the number of gestures used for more than one event increased over the sessions for three of the four children. To explain this result, they propose that the children move from a state in which they consider each of their gestures only in relation to the event it refers to, to one in which they consider their gestures in relation to each other as well. Only when a child does, Goldin-Meadow, et al. (1995) claim, does he notice that a particular gesture component (e.g., a particular handshape) recurs across gestures with some aspect of their meanings in common (in this case, presumably the shape of the object).[15] At that point, these recurring gesture components can "be separated from the wholes and treated as component parts", i.e. used productively. The effect is that very similar events (similar with respect to the semantic dimensions encoded by *both* gesture components) will be described with the same gesture.

This is an interesting theory, but several questions remain. For example, particular gesture components are used in gestures for different items from the beginning of the study. What makes the children suddenly notice recurrences that have been present all along? Further, in the beginning of the study, wouldn't a child have extended a particular gesture to a second item if that item had been similar enough to the first? If a child wanted to refer to an event with the same shape and motion as an event already referred to, and his other gestures are iconic, why would this gesture be any different? Unfortunately, this possibility was not tested.

What else might explain this result that gestures are increasingly extended to more than one event? It's possible that the children are categorizing more over sessions. For example, if a child begins to ignore the difference between being 1" wide and being 2" wide, characterizing each as simply narrow, then items that differ only slightly will come to be described with the same gesture. Thus, this homesign work suggests that a set of iconic signs can become systematic through categorization.

Unfortunately, Goldin-Meadow, et al. (1995) do not compare the systematicity of the children's gesture systems in the beginning of the study to that at the end. Although they examine how the gestures change in other ways, when addressing systematicity, they consider each child's system as if it were stable across the sessions.

[15] They call the regularities (the specific gesture components recurring in gestures for similar meanings) "haphazard", but this seems unlikely, given that the gestures are motivated. If the gestures for two different items include the same handshape component, it could be because the two items have similar shapes.

2.5 Summary

To summarize, previous work has demonstrated how novel signs can emerge and then evolve to become arbitrary, but does not measure systematic compositionality. Work on systematic compositionality proposes a mechanism for and measure of the evolution of systematic compositionality, but starts with arbitrary signs (i.e. does not address how arbitrariness emerges). Thus, there is a gap in the literature: it's not clear how the systematic re-use of arbitrary elements emerges in novel communication systems.

Chapter 3

Hypothesis

This thesis explores how the systematic re-use of arbitrary elements arises in a language. In the previous chapter, we surveyed the literature addressing this question. We saw that the signs people create are often motivated and that they can become arbitrary through interaction, but we did not learn much about how systematically compositional these novel signs are. We also saw how systematic compositionality can evolve in an artificial set of arbitrary signs in which both the set of signals and the set of meanings is structured, but this work did not directly address how novel signs might get to this stage. Thus, previous work has not painted a complete picture of how the systematic re-use of arbitrary elements evolves in novel sign systems. In this chapter, we offer an original hypothesis to fill this gap.

3.1 The parallel theory of the emergence of the systematic re-use of arbitrary elements

Of course, the simple combination of these two lines of research could produce the systematic re-use of arbitrary elements: first motivated signs evolve to become arbitrary, and then systematic compositionality evolves in that set of arbitrary signs. We'll call this the *serial theory* of the emergence of the systematic re-use of arbitrary elements.

We propose another possibility: the *parallel theory* of the emergence of the systematic re-use of arbitrary elements. In it, the emergence of arbitrariness and the emergence of systematic compositionality are not necessarily serial. Rather, novel signs make two potentially overlapping transitions: one to arbitrariness and the other to systematic compositionality.

How might this happen? There are two cases to be made here. First, systematic compositionality can emerge before signs have become completely arbitrary. Second, systematic compositionality can be maintained as signs become arbitrary. We argue these in turn now.

3.2 Systematic compositionality can emerge before signs have become completely arbitrary.

Let's first recall that novel signs are likely to be motivated. This is intuitive – if someone produces signals in order to convey meanings, they will choose signals that have some connection to those meanings. Otherwise, how can they expect anyone to recover their meanings? Further, the experimental work reviewed in Chapter 2 demonstrates that novel signs indeed tend to be motivated.

3.2.1 Due to chance

We know that systematic compositionality can emerge in a set of arbitrary signs, as long as the signals share elements with each other, the meanings share elements with each other, and it's possible for similar signals to map to similar meanings. (Kirby, et al., 2008). (Specifically, someone might notice a chance recurrence between a signal element and meaning element, e.g., that the names for two red things both include "ki." They might then use the mapping when speaking about a similar thing, thereby increasing the chance that someone else notices the regularity and does the same, and so on.) Although this was shown to hold for a set of arbitrary signs, there's no reason to believe that motivated signs (or signs falling anywhere on the motivated – arbitrary continuum) won't meet these conditions as well.

3.2.2 Due to motivatedness

The motivatedness of novel signs offers additional opportunities for systematic compositionality to emerge.

Immediate systematic compositionality due to motivatedness. A set of motivated signs might be systematically compositional from its birth. If someone wants to refer to two items with something in common (e.g., two edible items), it's at least plausible that they would produce similar signals for them (e.g., signals each involving a chewing motion of the mouth). Thus, a set of signs produced by this person could exhibit systematic compositionality, in which signals that share elements map to meanings that share elements. We introduced this relationship between motivatedness and systematic compositionality in Chapter 1.

Increasing systematic compositionality in motivated signs. Whether or not a set of motivated signs exhibits systematic compositionality immediately, i.e. from its birth, systematic compositionality can easily increase in a set of motivated signs.

This could be due to categorization. People might increasingly categorize semantic features, ignoring small differences in meanings (e.g., that one item is 1" wide and another 2"). In this case, in the initial sign system, few signals would share elements because even small differences in meaning would be represented in the signal. In later systems, more signals would share elements because these small differences would be ignored. Similarly, one might produce less careful signals. This is something like categorizing *signal* features. Imagine some buildings that typically take different shapes. For example, a typical fire station is not as tall as a typical hospital. Imagine also that someone is communicating by drawing. When initially drawing each building, they might aim to represent the shape of each well. But after drawing each several times, perhaps to reduce their production effort, they might simply draw a square (along with some distinguishing feature) for each. This again increases the systematic compositionality in the set of signs because the signals for buildings now share an element.

It's also possible that a set of motivated signs becomes more systematically compositional in a less gradual manner, i.e. by sudden replacement of the element of a given signal with a different one. Since many different signals (and signal elements) can have an inherent connection to a given meaning (as we argued in Chapter 1), people might be willing to replace one sign element with an equally motivated one that makes the set of signs more systematic. For example, someone might draw a school as simply a square (to indicate that it's a building) with a book (to indicate that it relates to primary education). Although they may have consistently drawn another item that relates to primary education (say, teacher) in an unrelated way (say, by drawing a chalkboard next to a stick figure), they may decide to replace the element in that drawing that indicates that it relates to primary education (the chalkboard) with the element used in the drawing for school that indicates this (the book). The set of signs is now more systematic because these two signals for similar items now share an element.

3.2.3 Due to interaction history

There is at least one more way systematic compositionality can emerge in novel signs: as a consequence of communicating with the same people about similar items. In this scenario, the first signal produced for a meaning would be highly motivated, but a subsequent signal produced for a similar meaning would be constrained not (only) by motivatedness but (also) by this first sign. Let's consider drawings of items relating to primary education as an example. The first time someone draws school for someone else, that drawing will be highly motivated. However, a later drawing of school bus between these two people may not be as motivated as if it were drawn between new partners. Rather, their drawing of school bus could reference their earlier drawing of school. We are suggesting here that, in the same way that communication partners converge on what signal to produce for a given meaning (Garrod, et al., 2007), they may also converge on what signal *element* to produce for a given meaning *element*. If they do, their set of signs will exhibit systematic compositionality.

3.2.4 Combinations of factors

Here we are arguing that systematic compositionality can emerge in novel, motivated signs, i.e. in signs that have not reached complete arbitrariness. This could happen in any of the ways just described, or even as a result of the combination of those factors. For example, both motivatedness and interaction history could simultaneously constrain what signal is produced for a meaning related to one already expressed. If a previously used signal element also has a strong inherent connection to the current meaning, it may be more likely to be used. For example, if someone draws teacher as a stick figure standing next to a school, he may have good reason to draw school bus as a vehicle next to a school. In this case, both motivatedness and interaction history may provide pressures to draw a school to signal primary education items. If, however, someone drew teacher as a stick figure next to a chalkboard, they may not be as willing to draw school bus as a vehicle next to a chalkboard. Despite the pressure from interaction history to draw a chalkboard to signal primary education items, the inherent connection between chalkboards and school buses might be too weak. (An iconic drawing of school bus is unlikely to include a drawing of a chalkboard.)

The relative strengths of these pressures may change over time. For example, the history of interaction may replace motivatedness as the main constraint on the signals over time. (Garrod, et al., 2007)

3.3 Systematic compositionality can be maintained as signs become arbitrary.

We are providing support for the possibility of the parallel emergence of arbitrariness and systematic compositionality in novel signs. In the previous section, we showed how systematic compositionality can emerge before signs have become arbitrary. We turn now to explaining how signs can become arbitrary without abandoning this systematic compositionality. We can approach this issue from two angles. First, as signs become arbitrary, how can systematic compositionality be maintained? Second, how can motivated systematic compositionality become arbitrary systematic compositionality?

3.3.1 As signs become arbitrary, how can systematic compositionality be maintained?

Garrod, et al. (2007) showed that signs can become arbitrary as the signals produced for frequent meanings simplify. They also suggested that the iconic components of a signal can become arbitrary while the whole signal retains its semantic complexity. In other words, sign elements - not just whole signs – could become arbitrary.

This scenario is plausible because sign elements shared across multiple signs will tend to be used more often than any particular whole sign. If interaction leads to simplification, signal elements should simplify more quickly than complex signals containing them. Imagine, for example, a child who communicates with his caretakers by using gesture. This child refers to the concept of tooth-brushing relatively often but in different contexts. For example, she communicates that she has already brushed her teeth, that brushing her teeth is boring, that her brother needs to brush his teeth, etc. We can imagine that the way she gestures tooth-brushing would initially be complex and motivated but then simplify to arbitrariness because she refers to tooth-brushing often. Any one of the more complex meanings involving tooth-brushing, however, would be expressed less often, and so how she gestures it would be less likely to simplify to arbitrariness.

3.3.2 How can motivated systematic compositionality become arbitrary systematic compositionality?

Another way to look at how a set of signs can become arbitrary while maintaining its systematic compositionality is to explore how *motivated* systematic compositionality can morph into *arbitrary* systematic compositionality. In the previous section, sign elements (mappings between signal elements and meaning elements) became more arbitrary because signal elements simplified, and this weakened the connection between signal element and meaning element. A sign element can also become arbitrary when the meaning of the signal element changes.

We first note that the issue of how motivated systematic compositionality can morph into arbitrary systematic compositionality is not trivial. Kalish, et al. (2007) demonstrated how difficult it is to replace a prior mapping with a new one. Thus, it may be difficult to change an inherent or natural connection between a signal and a meaning.

Yet we have already suggested one way the meaning of a signal element could change in a way that makes the sign element more arbitrary: the use of a signal element could be extended to other contexts. For example, if someone comes to use a representation of a particular building to refer to buildings in general, the connection between signal and meaning is weaker. Meanings of signal elements could change in other, similar ways. One could imagine the ways in which signals could come to map to less perceptual and more abstract or functional meanings. For example, the representation of a car could come to mean transportation in general and the representation of a wedding ring could come to mean love. In all of these cases, the sign elements have become more arbitrary.

3.3.3 Some degree of arbitrariness allows for (more) systematic compositionality.

Once sign elements have become somewhat arbitrary, systematic compositionality might further increase. For one, once signal elements relate less transparently to their original meanings, people may be more willing to extend them to other meanings. For example, if one's gesture for a particular building (say, movie theatre) comes to look less like that particular building, it may be easier for that gesture to refer to buildings in general. Similarly, if people prefer a consistent way of referring to things, less obviously motivated signs may allow them to change their other signs (or create new ones) to be more systematically compositional.

This increase in systematic compositionality as a result of the increasing arbitrariness of sign elements could happen during comprehension as well. Perhaps once sign elements have become somewhat arbitrary, someone realizes that not *all* signs should be interpreted as motivated, and stops interpreting *any* of them that way. The distinction between the existence of motivatedness in a language and people's awareness of it is key here. We suggest not that sign elements need to become completely arbitrary to allow systematic compositionality to increase, but rather that people need to be able to overlook (or not notice) any residual motivatedness. In this way, motivatedness may stop playing a role in the language before it completely disappears from it.

3.4 The co-evolution of arbitrariness and systematic compositionality

We have presented several scenarios in which novel, motivated signs could immediately exhibit or quickly develop systematic compositionality. We then suggested ways the sign elements could become arbitrary. To be clear, we are not implying that *first* systematic compositionality emerges and *then* arbitrariness emerges. As discussed, systematic compositionality could appear immediately, as a result of motivatedness, or could develop quickly. Each of the scenarios presented for the latter involves a move away from motivatedness and towards arbitrariness. That is, signs become more arbitrary and more systematically compositional in one step. Further, although systematic compositionality can emerge in novel, motivated signs, once sign elements have become somewhat arbitrary, systematic compositionality could further increase.

3.5 Summary

We've presented a parallel theory of the systematic re-use of arbitrary elements, arguing that systematic compositionality and arbitrariness are likely to have arisen in a parallel manner. Systematic compositionality can emerge before complete arbitrariness, and it can be maintained or increase as sign elements become arbitrary. In the next chapter, we begin to report a series of experiments that probe how the systematic re-use of arbitrary elements evolves in novel sign systems.

Chapter 4

Experiment 1

This thesis addresses how the systematic re-use of arbitrary elements can arise in a language. Thus far, we have established that previous work has not painted a complete picture of this – work exploring novel signs has shown that they are often motivated and then become arbitrary but has not directly addressed systematic compositionality, while work on the evolution of systematic compositionality has assumed a set of arbitrary signs as its starting point. We presented the parallel theory of the emergence of the systematic re-use of arbitrary elements, under which novel, motivated signs make two potentially overlapping transitions: one to arbitrariness and the other to systematic compositionality. In this chapter, we begin to report a series of experiments that test this possibility.

4.1 Experimental paradigm

Ideally, one could observe some people who had never been exposed to language start to communicate with each other. They'd be in a natural environment, communicating about whatever they wanted whenever they wanted. One would observe how they communicated from their first utterances up through the point where their language was relatively stable.

This ideal situation isn't feasible, but we designed an experimental paradigm to elicit signs like those that would occur in it. We discuss some features of the experimental paradigm here. First we note that we conducted a wide range of pilot experiments in order to determine the general features of the paradigm as well as the specific details of the actual experiments run. We will mention these pilots where necessary to motivate our choices.

4.1.1 Participants

Access to people who have never been exposed to language is difficult. Instead, we used university students as experimental participants and prohibited them from using any communicative conventions they had learned. Participants in a pair were separated from each other and could communicate only via an online whiteboard shared by the two. This alone prevents participants from using their normal means of communication (e.g., speech, gesture, and body language). But the participants were also prohibited from writing or using any drawing conventions (such as drawing a thick cross to symbolize the medical field).

Students with a background in Linguistics were excluded from the experiment, as their explicit knowledge of how languages work could have influenced the signs they created.

4.1.2 Duration of experiment

It was not feasible to observe participants steadily for months. Instead, the experiment lasted about two hours. Based on their feedback, this is as long as participants can be expected to stay comfortable and concentrate.

4.1.3 Task

We couldn't let participants communicate about whatever they wanted whenever they wanted because we had to ensure that at least some meanings were expressed many times during the two-hour experiment (so that we could see how signs changed over time). The first limitation we imposed to address this was restricting the communication to reference (to either tangible items or activities/situations). Reference is a fundamental function of communication, and so was the obvious place to start. Thus, our pairs of participants played a simple reference game. Each trial, one participant saw an item (a word, such as "teacher") and drew on a whiteboard, and the other participant guessed what the item was.

Note that the task of the participants was communication, rather than learning or imitating others' signs (as, for example, in Kirby, et al. (2008)). While there's surely a strong pressure to copy signs from others once a language is established, it's not clear how strong that pressure is from the very first signs – in the earliest stages of a language, people may choose to create their own, motivated signs rather than copy others' signs in order to communicate.

4.1.4 Number, frequency, and order of items

We also limited the number of items the participants express (instead of, say, putting them in some sort of game environment and letting them refer to anything). We used as many items as possible while still ensuring that each occurred often enough to allow us to study how the signals for each changed.

This might have introduced some artificial advantages in communicating. First, participants might have been able to narrow down what the current item was, e.g., by keeping track of what had already occurred. So that they couldn't, the items occurred with varying frequencies and in random order in the experiment. The different frequencies of the items also better matches real life, in which people communicate about some things more often than others.

Another artificial advantage that could stem from limiting the number of items involves participants realizing that the set of items is closed. In real life, people are free to refer to new items at any point in time. This means that, when someone utters something, their conversational partner can't assume they're referring to something the two have talked about before. Participants in pilots of Experiment 1 who knew their set of items was closed reported behaving differently than participants who did not realize this. In particular, participants who knew their set of items was closed were more willing to draw something for an item that was different from what they'd drawn for that item in the past. To eliminate this advantage, the participants had to communicate a brand new item every so often throughout the entire experiment.

4.1.5 Who communicates what when

In natural settings, it's unusual for someone to explicitly determine whose turn it is to speak. This is difficult to avoid in controlled experiments, though.

To make the experiment more like real life, we had participants switch roles (drawing or guessing) frequently. One desired effect of this was that some of the items were first communicated by one participant and the remainder of the items were first communicated by the other participant.[16] Participants in pilot experiments (in which the set of items seemed open, as in Experiment 1) reported sticking with their first drawing for each item – not adding to it or replacing parts of it during the game. If this is the case, the first person to draw each item effectively chooses how that item will be drawn throughout the rest of the game. This experiment was designed to address how people *co*-create a communication system, so we had each participant create some of the signs.

Frequent role-swapping and the random appearance of the items means that we didn't control when each participant drew each item. One disadvantage to this is that it's less straightforward to identify a participant's drawing of an item at any particular point in time in the game, such as at the beginning or end of the game, because a participant may happen not to draw a certain item for many trials. For example, we may not know what one participant would have drawn for "tractor" at the beginning of his game because, by chance, he didn't get to draw "tractor" for several rounds.

4.1.6 Semantics of the items

Items were chosen to share semantic features. Since there are a limited number of items, and we are exploring systematic compositionality, the items had to be carefully selected to share salient semantic features. Otherwise, systematically

---

[16] This is in contrast to Garrod, et al. (2007), where, when Director and Matcher roles were swapped, they were swapped only after all the items had been drawn.

compositional sets of signs, where items that share features are mapped to signals that share elements, aren't possible.

One might be concerned that the fact that all of the items can be seen as similar to each other produced an artificial pressure for systematic compositionality to emerge in the sets of signs. However, in post-experiment interviews, many participants reported not even noticing the overall organization. While they certainly know that "firefighter" and "teacher" are both a kind of person, they didn't realize that *all* of their items could be arranged according to just two dimensions.

Real items, not constructed stimuli. Although the items were chosen to have semantic overlap, we could not *guarantee* semantic overlap because we used real items as stimuli[17]. For example, it may not have been salient to all the participants that "tractor" has something in common with "school bus". This decision to use real items contrasts with previous work. For example, each of Kirby, et al. (2008)'s stimulus images was one of three colours, had one of three shapes, had one of three motions, and had no other semantic features. Likewise, each item in Selten and Warglien (2007)'s experiment was one of a few shapes, with one of a few inserts (or none), with one of a few colours (or not coloured). Our decision to use real items was important, for this thesis explores systematic compositionality in natural sign systems. One step in creating a systematically compositional set of signs is noticing the semantic overlap between the items. Thus, using artificial stimuli in which the semantic overlap may be unnaturally salient skips a possibly formative stage. One pilot experiment provided an example of collapsing unimportant distinctions, such as the different shapes of particular vehicles.

Figure 1 shows the initial and final drawings of "fire engine" and "tractor" produced in this pilot. Notice how, by the end of the game, details of the vehicles (such as their respective shapes) are left out. This collapsing of fine distinctions (so that, for example, fire engine and tractor can be seen as simply a vehicle that relates to fire

[17] Technically, our stimuli are words referring to real items.

and a vehicle that relates to farming, respectively) is a barrier to systematic compositionality that sets of novel, motivated signs may have to break through.

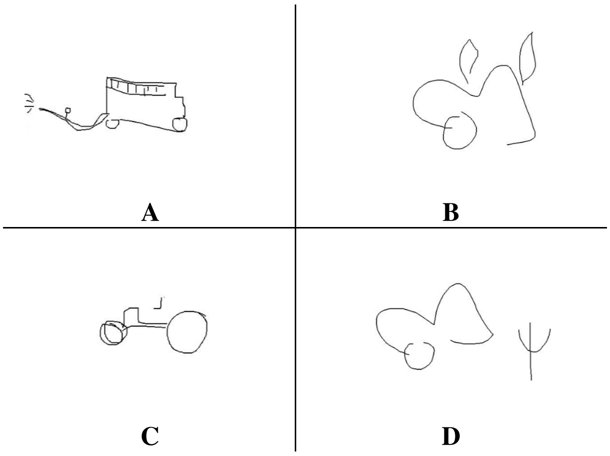


Figure 1. Initial drawing of "fire engine" (A), final drawing of "fire engine" (B), initial drawing of "tractor" (C), and final drawing of "tractor" (D) from a pilot experiment. By the end of the game, the pair no longer represents the different shapes of the vehicles.

Some constructed sets of stimuli may have another artificial pressure towards systematic compositionality: the items lack unique features. For example, "tractor" appears to be the only item in our stimulus set that has two different-sized wheels. If a pair of participants drew "tractor" as just two different-sized wheels, their set of drawings could be less systematically compositional because no part of the drawing for "tractor" will occur in drawings of other vehicles or in drawings of other items that relate to farming. When each stimulus item is nothing more than a set of feature values shared with other stimuli, representations of them could be almost guaranteed to be systematically compositional – how else could one uniquely identify that item? We avoided this problem by using real items as stimuli.

No direct pressure for systematicity. Since we are exploring systematic compositionality in natural signs, it was important that the experiment introduced no direct pressure to create systematically compositional sets of signs. The set of items is very structured, so this was tricky. The first way we ensured this was by never presenting the full set of items to the participant - the participants never learn what items they will communicate about; the items simply appear in the game. For example, as we explain below, participants don't guess by choosing from a list of possible items. Rather, they guess without any hints by typing into a chat window. This means the participants didn't have a chance to spot the semantic structure of the set of items before they created signals for them. In contrast, Figure 2 shows the first signs produced for items relating to primary education in a pilot experiment in which the participants studied their items, which were presented in a table with row labels (e.g., "person") and column labels (e.g., "primary education"), before their game began. Right from the beginning of this game, the participants represented each item as a combination of the two feature values represented in the table (viz. as a person, building, or vehicle relating to primary education).

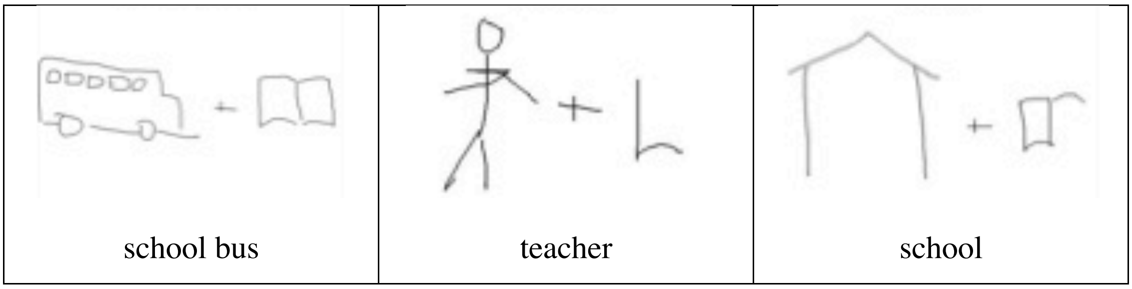


Figure 2. First drawings of "school bus", "teacher", and "school" from a pilot experiment in which participants studied their items, organized in a labelled table, before playing their game.

The second way we ensured that there was no direct pressure to create systematically compositional sets of signs was by introducing a prize draw for the top-performing pairs of participants. The prize was large enough (about double the compensation for participation) that participants could be expected to care more about their success in the game than making their set of signs systematically compositional, even if they believed that the latter was what the experimenter wanted from them.

4.1.7 Communication medium

In the ideal observation setting described above, people might try to communicate using their voices, their motions, their gestures, etc. In each of these, it may be more difficult to produce motivated signs than it is to produce motivated signs on a whiteboard. To make the communication medium in our experiment more restrictive, participants drew with a mouse (with which is harder to draw detail than with a stylus pen) and could only draw in black ink. Participants sometimes complained that it was difficult to draw with the mouse.

The participants were not allowed to use their normal means of communication (mainly, speech) because the experiment was meant to inform how people communicate in the absence of conventions. Unfortunately, the graphical medium has its own communicative conventions. Thus, the instructions explicitly prohibited the use of symbols and drawing conventions in the game: letters, numbers, punctuation (?, !, etc.), mathematical signs (+, =, <, etc.), pointing with arrows, universal graphical signs (the cross for first aid, a skull and cross-bones for danger, etc.), and drawing conventions (a heart to represent love, + over the eyes to show that someone is dead, etc.).

Participants were also prohibited from marking prominence in any *conventional* way. For example, the instructions explicitly prohibited drawing arrows on the whiteboard. But marking prominence may be a natural, pervasive ability, so we didn't explicitly prohibit marking prominence in general (e.g., by circling or tracing over a part of a drawing).

One particular thing to note about our communication medium is that the signals aren't necessarily combinatorial, i.e. they don't necessarily share elements with each other. This contrasts with some previous work where the participants created signals from a limited set of letters or syllables. Since we're interested in systematic compositionality (the re-use of signal elements in signals for similar meanings), we did not want to start from combinatoriality (the re-use of signal elements). In other words, we did not assume combinatoriality, but rather observed what kind of signals the participants created.

Note that this communication medium allows the signs created using it to be motivated. This is important for at least two reasons. First, some models of the evolution of systematic compositionality take as their starting point a set of signs that is arbitrary but whose signals are combinatorial. Yet, we saw in Chapter 2 that novel signs are often motivated. They can become arbitrary as the signals lose complexity, but as signals become less complex, the set of signals may become less combinatorial – as each signal simplifies, it may be less likely to share an element with another signal. Thus, it may not be straightforward for a set of novel signs to become arbitrary while its signals retain combinatoriality. Our experiment does not assume such a set of signs as its starting point. Rather, participants create the initial signs.

Allowing motivatedness is important for another reason: in previous work in which the sets of signs started off arbitrary, the only way to increase the expressiveness of the language (given a bottleneck) was to increase the systematic compositionality of the set. Systematic compositionality was found to be an adaptation to a bottleneck. But since novel signs are likely to be motivated – the features of their signals chosen precisely because they do convey meaning – this experimental design may have produced a somewhat artificial pressure for systematic compositionality. In other words, when people can produce motivated signs, a bottleneck may not have the same effect.

While motivatedness was possible in our experiment, it wasn't forced. In the ideal setting described above, people aren't told anything about how to create signs. They would presumably be comfortable generating any signal that conveyed the given meaning. Likewise, in Experiment 1, the participants are never instructed to "draw the item". The participants are told they're going to communicate items to each other by drawing on a whiteboard and that the Drawer's task is to get the Guesser to guess the correct item.

4.1.8 Pressure to reduce production effort

When people refer to something often, their signal for that meaning shortens. (Clark & Wilkes-Gibbs, 1986) Since Experiment 1 necessarily condensed the timeframe of this natural process, we introduced a direct pressure to reduce production effort. We did this by encouraging participants to complete as many trials as possible in the allotted time (by awarding the top-performing pairs with an entry into a prize draw for extra cash). Instead, we could have rewarded participants for completing a predetermined number of trials quickly. But pilots showed – and the experiment later confirmed – that participants' speed varied widely. We would have gathered less data (viz. fewer drawings) if each pair completed only as many trials as we could expect *all* pairs to complete during the experiment.

4.1.9 Motivation to communicate successfully

In the ideal setting, people would communicate when they wanted to. Presumably, anytime they would communicate, they would care to do so successfully. We ensured that our participants were similarly motivated to communicate successfully by implementing a scoring system that rewarded correct guesses and penalized incorrect guesses. Crucially, as mentioned above, there was a prize draw for a good amount of cash (about double the compensation for participation) for top-performing participants. So that participants didn't get so frustrated by their poor or slow performance that they gave up on the game entirely, the *three* top-performing pairs were entered into a prize draw. This way, participants were more likely to feel like they still had a chance at the prize and stay motivated throughout the game.

We penalized incorrect guesses so that participants wouldn't try to skip difficult items. Otherwise, a clever Drawer could simply pause when he got an item he knew would take a lot of time to communicate, and the perceptive Guesser then might have guessed anything at all, just to get to the next trial.

Participants were allowed just one guess per trial. This is a bit artificial, but made analysis straightforward - whatever was drawn on the whiteboard was taken to be that participant's signal for the item (as opposed to the whiteboard including responses to the Guesser's guess).

4.2 Method

Given these general design considerations, Experiment 1 was designed.

4.2.1 Participants

24 University of Edinburgh students participated in exchange for £12. All were native British English speakers. Students with a background in Linguistics were

excluded. Participants who played together didn't know each other. As the advertisement called for native British English speakers, participants probably assumed they were playing with another native British English speaker.

4.2.2 Apparatus

Partners were seated in separate soundproof booths, each in front of a computer monitor, keyboard, and mouse. The experiment was run using the Pigeon software (Healey, Swoboda, & King, 2002), which presented the item each trial and provided the shared online whiteboard. Participants guessed and corrected their partners' guesses in a separate MSN Messenger chat window. Also on the screen was a timer counting down from two hours.

4.2.3 Stimuli

The items that were used as stimuli are shown in Figure 3. They were chosen to share salient semantic features; each item can be thought of as one of five entity types (such as person or building) that relates to one of ten themes (such as university education or agriculture).

There were 26 core items, each of which occurred multiple times per game, and 14 filler items, each of which occurred at most once per game. The participants played for two hours and were instructed to complete as many trials as possible. Pilots of the experiment indicated that each pair could be expected to complete at least 126 trials in the two hours. Thus, the frequency distribution of the items is based on blocks of 126 trials. Every 126 trials, four of the items (in the innermost box in Figure 3) appeared 8 times each, eleven of the items (in the middle box in Figure 3) appeared 5 times each, and eleven of the items (in the outermost box in Figure 3) appeared 3 times each.

teacher	school	teaching	school bus	classroom
firefighter	fire station	fire-fighting	fire engine	
professor	university	lecturing		lecture theatre
doctor	hospital	medical emergency	ambulance	emergency room
chef	restaurant	cooking		gourmet kitchen
farmer	barn	farming	tractor	
soldier	barracks	war	tank	
prisoner	jail	crime	police car	
chemist	pharmacy	prescription		
dentist	dental practice	root canal		

Figure 3. The items used as stimuli. Every 126 trials, each of the four items in the innermost box appeared 8 times, each item in the middle box appeared 5 times, and each item in the outermost box appeared 3 times. Every 126 trials, six of the unboxed (filler) items appear. Each of these unboxed items appears at most once per game.

Every 126 trials, six of the fourteen filler items (unboxed in Figure 3) occurred. Each filler item occurred at most once per game.

Within each block of 126 trials, the items occur in random order.

4.2.4 Rules of the Game

The full instructions to the game are included as Appendix A.

They instructed the participants not to use symbols or drawing conventions.

A team was allowed just one guess per trial.

A team won one point for every correct guess but lost one point for any incorrect guess or drawing that included a symbol or convention. The goal was to win as many points as possible in the two hours of play.

Participants from the three top-scoring teams were entered into a prize draw for an additional £20.

4.2.5 Procedure

Before the experiment started, participants read instructions. They did not see a list of the items they were about to communicate, nor know anything about the items.

Each trial, one participant was the Drawer and other was the Guesser. The Drawer saw an item (a word, such as “professor”) on his screen and was allowed to draw immediately. The Drawer drew with a mouse, had only black ink, and could not erase anything. The Guesser saw everything the Drawer drew immediately, on her screen. The Guesser did not see the Drawer's mouse movements when he was not drawing, and could not draw herself.

When she was ready, the Guesser guessed by typing into a chat window. The Drawer stopped drawing immediately and typed the item into the chat window, whether or not the guess was correct. Nothing but the guess (typed by the Guesser) and the exact item (typed by the Drawer) were allowed in the chat window. The chat window was sized to show only the previous three guesses or items, and participants were not allowed to scroll up to see earlier items. Once the Guesser read what the item was in the chat window, she hit a “Next Item” button. Both participants hit an “OK” button to start the next trial.

Every six trials, the participants switched Drawer and Guesser roles. Both participants had to hit a button in order to advance to the next round of six trials. Participants were encouraged to take any breaks they needed in between rounds.

The participants played for two hours.

4.3 Results

4.3.1 Success

Game score. Game scores varied widely – ranging from 66 to 260 points, with an average of 150.67 points (SD = 59.21).

Pairs received one point for every correctly guessed trial but lost one point for every trial which was incorrectly guessed or in which a symbol or drawing convention was used. Thus, success in the game depended on how many trials were completed, how many trials were incorrect, and how many trials included symbols or drawing conventions. Let's look at each of these in turn.

Trials completed. Pairs completed between 138 and 330 trials in the game. The average was 232.5 trials (SD = 58.91). Game score was strongly correlated with the number of trials completed (Spearman's $\rho = 0.61$, $p < 0.05$).

Incorrect guesses. Pairs made between 20 and 70 incorrect guesses in their games (M = 36.17, SD = 15.34). There were an average of 37 unique items in each game (SD = 2.57), and each might be expected to be guessed incorrectly the first time it appears in the game. On average, 72% of the incorrect guesses in a game were made on items making their first appearance in the game (SD = 16.45%). There is no significant correlation between game score and number of incorrect guesses.

Symbol use. Pairs generally followed the instructions not to use symbols or drawing conventions. In fact, seven of the 12 pairs used no symbols or drawing conventions at all during their games. The remaining five pairs used a convention in 1, 5, 5, 26, and 32 trials of their games. Interestingly, only two conventions were used by the five pairs of participants who used any: (1) the first aid cross, used in drawing of medical items, and (2) speech bubbles, used in drawings of primary education and university education items, especially "teaching" and "lecturing". There is no significant correlation between game score and the number of trials in which symbols or drawing conventions were used.

4.3.2 Systematic Compositionality

Figure 4 shows one of the sign systems that emerged from this game. Each image shows what was drawn on the whiteboard for that item the last time it occurred in the game. Items in italics were drawn by one participant, while items in the normal typeface were drawn by the other. The signs are arranged according to the semantic features of the items, not by the chronological order of the trials. Items in the same row and items in the same column share a semantic feature. For example, the first column contains this pair's drawings for various people.

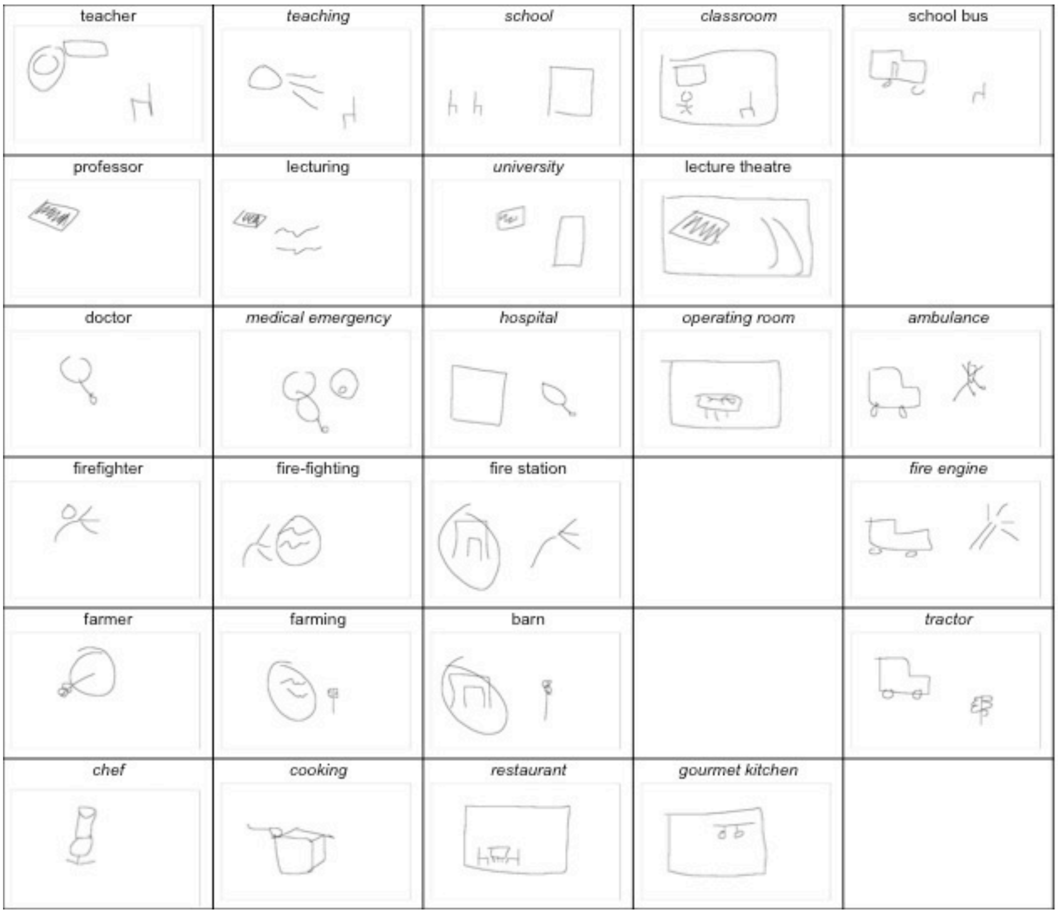


Figure 4. One set of signs that emerged from Experiment 1. Each image shows what was drawn on the whiteboard for that item on the last occurrence of that item in the game. Signs are arranged according to the semantic features of the items, not by chronological order of the trials. Italics vs. normal typeface distinguish which participant was the Drawer that trial. The set is highly systematically compositional, in that signals in many of the rows and columns share an element.

Notice how systematically compositional the set of signs is: the drawings in many of the rows and columns share an element. For example, the drawings for items relating to university education (in the second row) each have a filled-in diamond. As another example, four of the drawings for activities/situations (in the second column) have a row of squiggly lines.

While the systematic compositionality in the set of signs in Figure 4 is striking, devising a way to measure it was not straightforward.

4.3.2.1 Previous Measures of Systematic Compositionality

Previous work has measured systematic compositionality using methods not directly transferrable to our data. For example, Kirby, et al. (2008) measured systematicity as a Pairwise Distance Correlation (PDC) - the correlation of similarity between signals with similarity between their corresponding meanings. They then calculated the extent to which this correlation differs from that we would expect to see if the signals were randomly assigned to meanings. Because signals were character strings and meanings were sets of discrete feature values, similarity could be calculated as edit distance - the smallest number of insertions, replacements, and deletions required to transform one into the other. Calculating edit distance between the signals or between the meanings in Experiment 1 would not be as straightforward, because neither the signals nor the meanings are necessarily composed of discrete elements. It might be done using human judgment, but the number of required judgments would be very high.

RegMap (Tamariz & Smith, 2008) measures the degree of regularity in the mappings from signals to meanings (or vice versa) and does not require any similarity measures between signals or between meanings. For each possible mapping from signal element to meaning element (or vice versa), one compares the number of examples of that mapping with the number of counterexamples to it. Unfortunately, we would run into a problem similar to the one with Pairwise Distance Correlation using this method - because neither the signals nor the meanings in Experiment 1 are necessarily composed of discrete elements, counting examples and counterexamples of element mappings would require human judgment – a coder would have to break every signal (drawing) down into elements. This is unnecessarily labour-intensive for our current purposes; we don't need to know about the idiomatic signs in a set. Further, we don't need a coder to consider every possible mapping between signal element and meaning element; we just need him to pick out the element mappings he thinks the original participants were using.

Goldin-Meadow, et al. (1995) faced the problem of identifying signal elements as well, when they measured regularities within sets of homesign gestures. They solved it by defining signal elements themselves, as categories of feature values. Then two independent coders characterized each gesture using these experimenter-defined categories. Finally, the experimenters calculated the proportion of each child's gesture tokens that fit the system they identified. The proportions are between 70% and 95%. They say these fits are "quite high", but don't provide any baseline.

4.3.2.2 Current Measure of Systematic Compositionality

We draw from all of these approaches.

Here, a set of signs consists of one signal for each of the 26 core items. For example, a pair's Final set of signs consists of the signal produced for each item the last time that item occurred in their game. The experimenter coded each set for systematic compositionality. Each set of drawings was printed on a page in a table, organized so that rows and columns contain drawings for similar items (as in Figure 4). The coder examined each row and each column for any element shared among two or more drawings. If there was a shared element, the coder marked which of the drawings in that row or column included it. The coding instructions are included as Appendix B.

There are 26 drawings (one for each of the core items) and each drawing is inspected twice – once as a member of its row and once as a member of its column. Thus, each set of drawings can receive a total score of 52. The total score divided by 52 is our systematic compositionality score.

For each of the 12 games of the experiment, we coded the Initial sets of drawings (the first drawing of each item from that pair of participants) and the Final sets of drawings (the last drawing of each item).

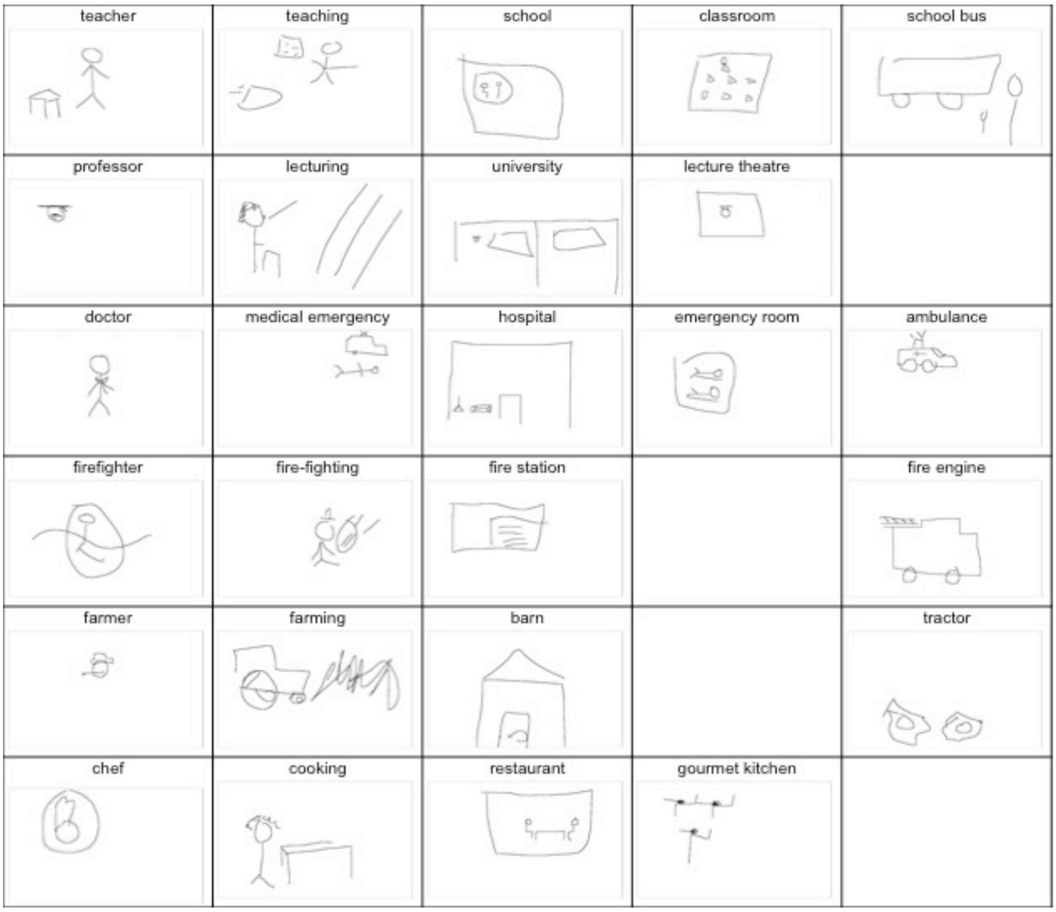


Figure 5. A Mixed Final set of signs. Each image shows what was drawn for the item the last time it appeared in a game, but the images within a row or column are drawn from different games. There is little systematic compositionality in the set – only rarely do drawings in a row or column share an element.

To determine baseline systematic compositionality, we constructed 12 sets each of two kinds of control sets: Mixed Initial and Mixed Final. The Mixed Initial sets were each composed of the initial drawings from *different* games of the experiment (i.e. from different pairs of participants). For each Mixed set, for each item (e.g., "teacher"), we choose at random which of the games the drawing would be from, with the restriction that the drawings in each row and each column would be from different games. The Mixed Final sets were constructed in an analogous manner. Figure 5 shows one Mixed Final set of signs. Thus we have four categories of sets of signs – Initial, Final, Mixed Initial, and Mixed Final – and each includes 12 sets of signs.

The coder marked these 48 sets in random order and blind, i.e. she didn't know which set she was coding.

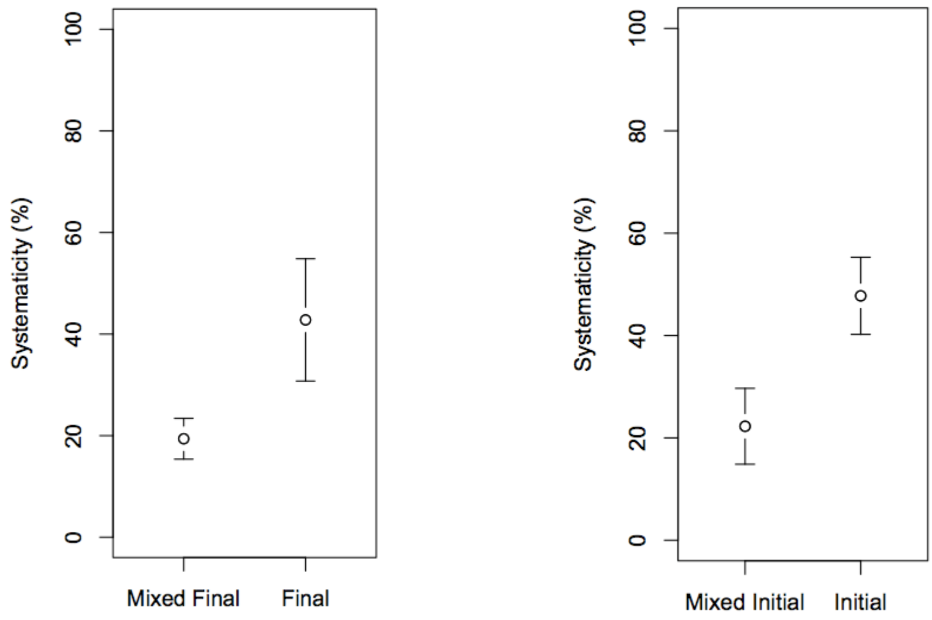
4.3.2.3 Reliability

To check whether the experimenter's coding was biased, a different coder – who had no role at all in the experiment – marked three randomly chosen sets of each category (12 sets in total) independently. Her scores were strongly correlated with those of the original coder (Spearman's $\rho = 0.82$, $p = 0.001$). (The sets of systematic compositionality scores failed standard tests for normality, and so non-parametric statistics were used here and in subsequent systematic compositionality analyses.)

4.3.2.4 Results

Figure 6 shows the mean systematic compositionality for each of the four categories of sets of signs.[18] ($M_{\text{Final}} = 42.79$, $SD = 18.95$; $M_{\text{Mixed Final}} = 19.39$, $SD = 6.32$; $M_{\text{Initial}} = 47.76$, $SD = 11.85$; $M_{\text{Mixed Initial}} = 22.28$, $SD = 11.67$) The first thing to notice is that Final sets of signs are more systematically compositional than Mixed Final sets of signs (Figure 6A). That is, pairs' last drawings of items systematically re-use drawing elements more than can be attributed to a tendency across pairs to do so – they are truly systematic. A Mann–Whitney U Test confirmed this ($p < 0.01$).

[18] One of the Mixed Initial sets of signs is an outlier, with a systematic compositionality score more than 1.5 times the interquartile range above the third quartile for Mixed Initial sets of signs. We chose not to exclude it from analysis because there appears to be no error in creation or measurement of that set; it's simply unusually systematic for Mixed Initial sets of signs.



A B

Figure 6. Mean systematicity (%) and confidence intervals (confidence level = 95%) for Initial, Mixed Initial, Final, and Mixed Final sets of signs. Final sets are more systematically compositional than Mixed Final sets and Initial sets are more systematically compositional than Mixed Initial sets.

How did Final sets of signs get to be systematically compositional? It turns out that Initial sets of signs are also truly systematically compositional; see Figure 6B. A Mann-Whitney U Test confirmed that the Initial sets of signs have significantly higher systematic compositionality scores than the Mixed Initial sets of signs ($p < 0.001$).

Further, sets of signs didn't become significantly more or less systematically compositional over time. A Wilcoxon Signed-Rank Test found no significant difference between the systematic compositionality scores of Initial and Final sets of signs ($p= 0.2662$).

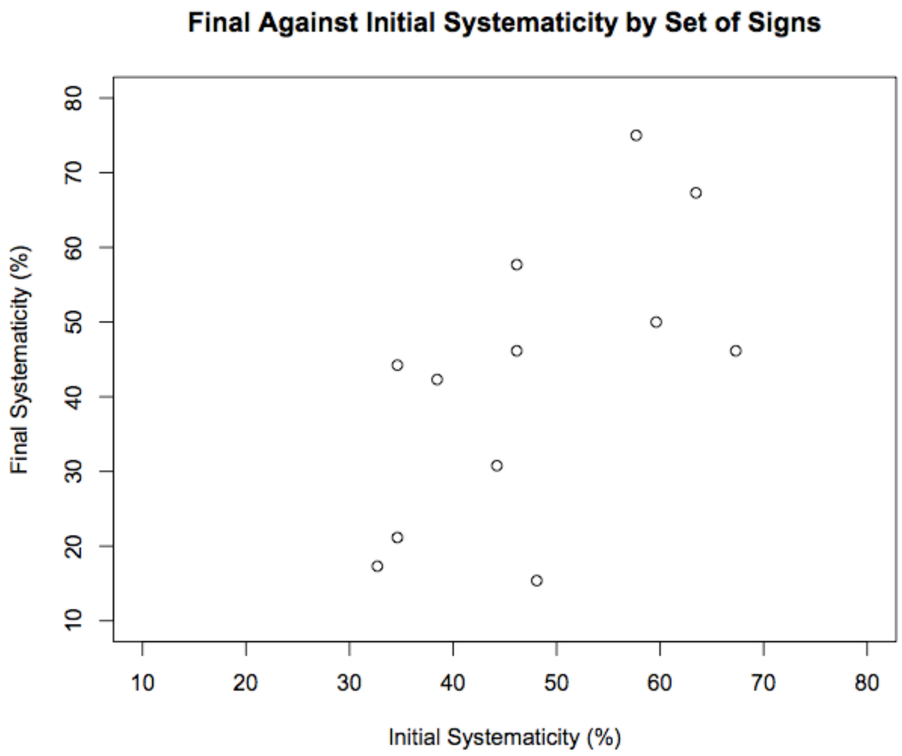


Figure 7. Scatterplot of Final against Initial Systematicity (%). Spearman's ρ equals 0.62 (p=0.03), indicating there is some correlation between the ranking of the Initial sets of signs by systematicity and that of the corresponding Final sets of signs.

What then is the relationship between the degree of systematic compositionality in a pair's Initial set of signs and the degree of systematic compositionality in their Final set of signs? Figure 7 is a scatter plot of each of the 12 pair's Initial systematic compositionality score against their Final systematic compositionality score. There is clearly a correlation between the Initial and Final systematic compositionality of a pair's signs. (Spearman's $\rho = 0.62$, $p < 0.05$).

4.3.3 Arbitrariness

As in the systematicity analyses, we examined the arbitrariness of the Initial and Final sets of signs from each pair. We hypothesized that, as in Garrod, et al. (2007), the arbitrariness of the signs increased through interaction.

4.3.3.1 Procedure

To measure how arbitrary the signs produced in Experiment 1 were, we followed Fay, et al. (2008) and had new participants guess what the signs meant.

12 University of Edinburgh students participated in exchange for chances in a prize draw for a £25 Amazon gift voucher. All were native British English speakers. The corpus was sampled once across participants; each sign was judged once. A mixed design was used: each participant was presented with the Initial drawings of each core item from one original game and the Final drawings of each core item from a different original game. The participants were assigned games at random.

The experiment was run over the Internet, and lasted approximately 15 minutes. Participants read about the basics of the original communication game and learned that the drawings they'd see would be from different games and different points in time during the games, in random order. Participants received an entry in the prize draw for each correct guess.

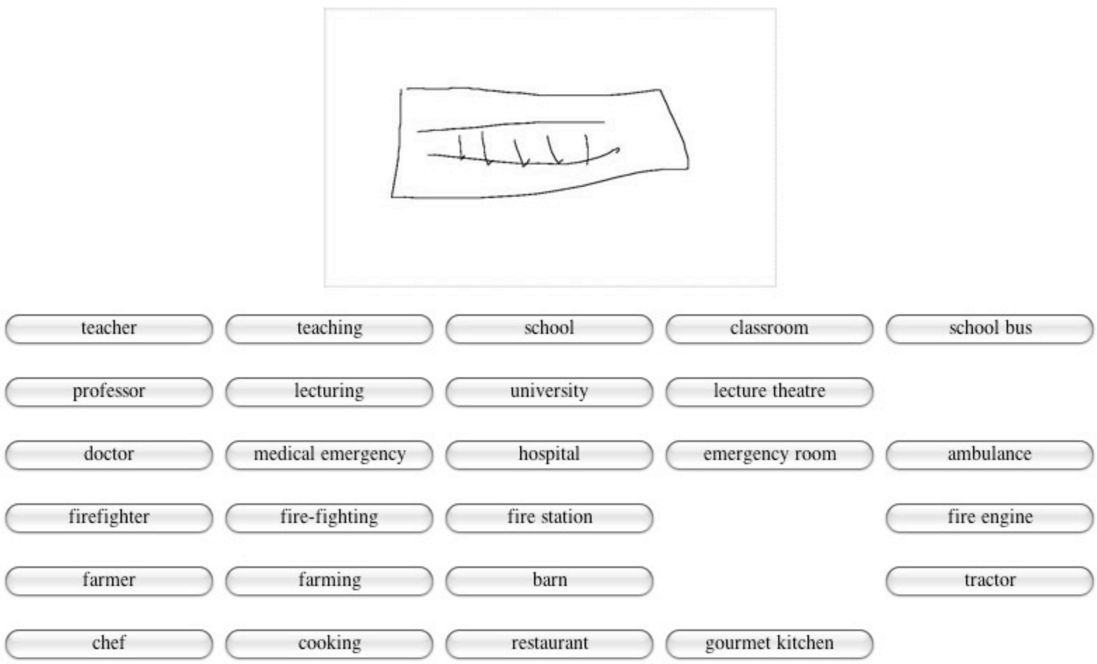


Figure 8. The screen for a participant making a transparency judgement.

Each trial, a participant saw a screenshot of the whiteboard at the end of the trial in the original game. He guessed the meaning of the image by clicking on one of 26 buttons, one for each possible item. Figure 8 shows what the screen looks like for an example trial.

4.3.3.2 Results

Figure 9 shows the mean identification rates (as proportions correct) for Initial and Final sets of signs. ($M_{Initial}$ = 64.08, SD = 12.37; M_{Final} = 45.42, SD = 6.86) Initial sets of signs were more accurately identified than Final sets of signs. A Mann–Whitney U Test confirmed this ($p < 0.001$). This suggests that the signs became more arbitrary – their signals related less transparently to their meanings – over the course of the games.

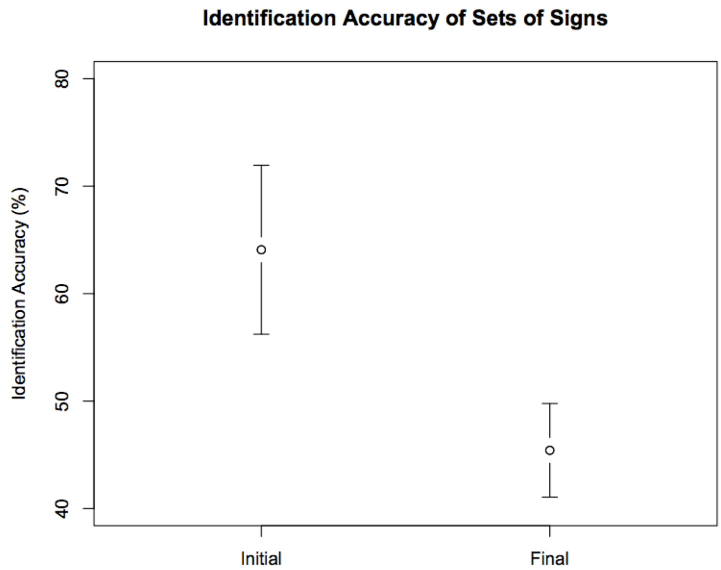


Figure 9. Mean identification accuracy (%) and confidence intervals (confidence level = 95%) for Initial and Final sets of signs. Initial sets of signs are more accurately identified than Final.

4.4 Observations

In Chapter 3, we suggested several ways systematic compositionality and arbitrariness could emerge in parallel. A qualitative exploration of the signs produced during Experiment 1 reveals that many of the possibilities we suggested were realized.

4.4.1 Systematic compositionality due to interaction history

The first part of our argument for the parallel theory of the emergence of the systematic re-use of arbitrary elements included reasons we could expect a set of novel signs to be systematically compositional from its genesis, or to straightforwardly become so. We suggested that this could be due to a combination of motivatedness and interaction history. We'll consider interaction history first.

Immediate systematic compositionality due to interaction history. Perhaps the most surprising result from Experiment 1 is that sets of signs are systematically compositional immediately, i.e. even the very first signs a pair creates share elements with each other. The elements are specific to the pair; sets of Initial signs drawn from *different* pairs exhibited less than half the systematic compositionality.

Recall that participants have no access to a list of items they will be communicating before the game; it's not that participants are designing systematically compositional sets of signs in advance. Further, there is no direct pressure for systematic compositionality in the game; each pair's goal is simply to communicate accurately and quickly.

What is behind this immediate systematic compositionality then? While the experiment was not designed to distinguish the effects of motivatedness from those of interaction history, there is reason to believe that interaction history plays a larger role than motivatedness.

Systematic compositionality is measured on a *pair*'s set of signs, not an individual player's. Specifically, the Initial set for a pair consists of the drawing produced for each item the first time that item appeared in the game, regardless of which participant was the Drawer that trial. On average, half of the drawings in a set are produced by one participant in the pair and half are produced by the other.

For the systematic compositionality of a set of signs produced by two players *not* to relate to the interaction history shared by the two either (1) one player is solely responsible for the systematic compositionality in the set or (2) the two players create the same sign elements by chance (i.e. they would produce those sign elements regardless of partner).

It's unlikely that one player in each pair is solely responsible for the systematic compositionality in the set of signs produced by both him and his partner. The initial systematic compositionality scores average almost 50%. The one participant's subset of the signs would have to be perfectly systematically compositional for the whole set of signs to achieve this. Further, the Initial systematic compositionality scores reach as high as 67%. Any systematic compositionality score significantly higher than 50% cannot be explained by the self-systematicity of one of the players.

It's also unlikely that both players in a pair just happened to choose the same way to refer to the items. For a particular meaning, there are many different motivated ways to draw it. Consider the drawings of "farmer" in Figure 10, drawn from various games (of Experiment 1 and similar pilot experiments). Clearly, participants associate many different things with the concept of a farmer: various settings (rolling hills, sunshine, plowed land, fruit trees), nearby items (tractor, barn, hay), animals (cows, pigs, chickens, a hunting dog), items of clothing (overalls, a plaid shirt, a straw hat, Wellies), instruments (spade, pitchfork, walking stick), activities (chewing corn, plowing land), and so on. In fact, this wide variety of signals for a given meaning is what the low systematic compositionality of the Mixed Initial sets illustrates. Thus, the immediate systematic compositionality of a pair's set of signs is likely to be due to the interaction history shared by the two.

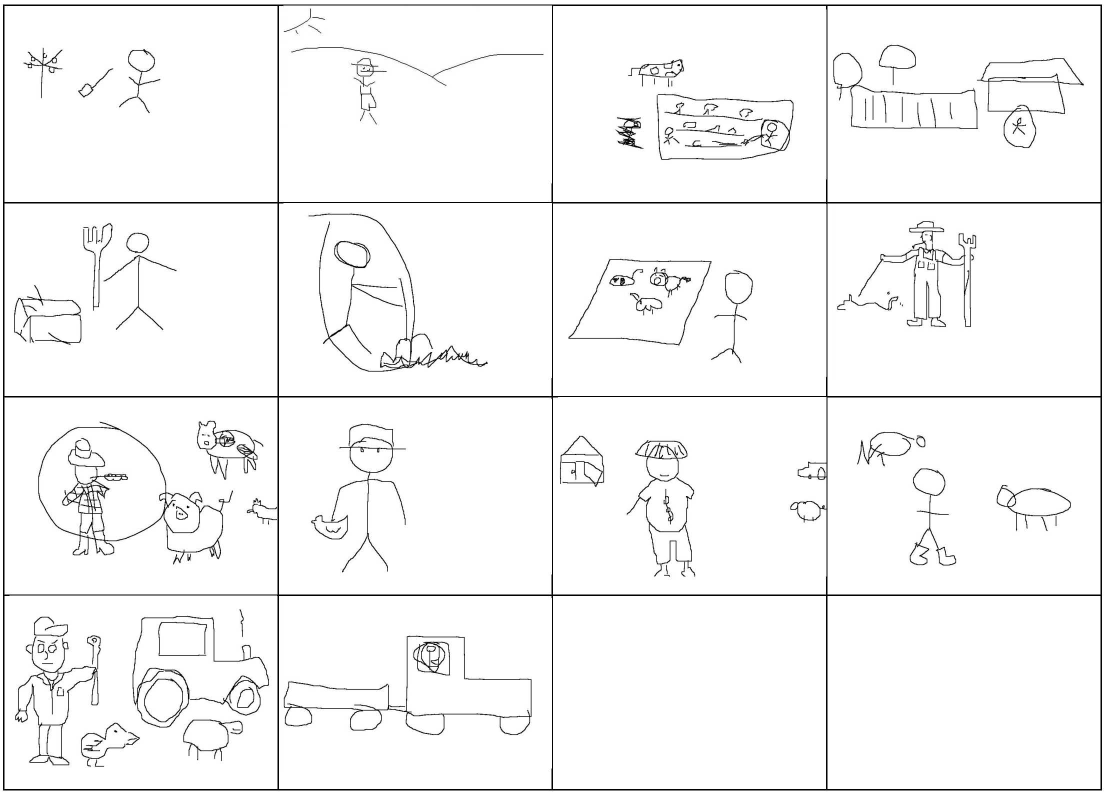


Figure 10. Drawings of "farmer" across games of Experiment 1 and similar pilot experiments. The concept of a farmer is associated with many different settings, animals, items of clothing, instruments, and activities.

To explore whether motivatedness or interaction history explains immediate systematic compositionality better, we can also find occasions in which they are at odds. Consider, for example, the drawings of "school bus" in Figure 11. For pairs A – D, "school bus" was the first primary education item to be drawn, so the drawings reflect what people draw for "school bus" when uninfluenced by previous drawings of primary education items – roughly, what motivated drawings of "school bus" look like. In contrast, Pair E drew "school bus" *after* they'd already drawn another primary education item: "teacher." Their drawing of "school bus" appears to re-use elements from their drawing of "teacher", i.e. be strongly influenced by signs in pair's shared interaction history. In particular, note that a chalkboard in a drawing for "school bus" is not strongly motivated – none of the uninfluenced drawings of "school bus" include a chalkboard. While one might expect that the first time someone draws "school bus" with their partner, they would draw it no differently than if they were drawing with a new partner, instead it appears that the history of interaction shared by two people influences even new signs they create for use with each other.

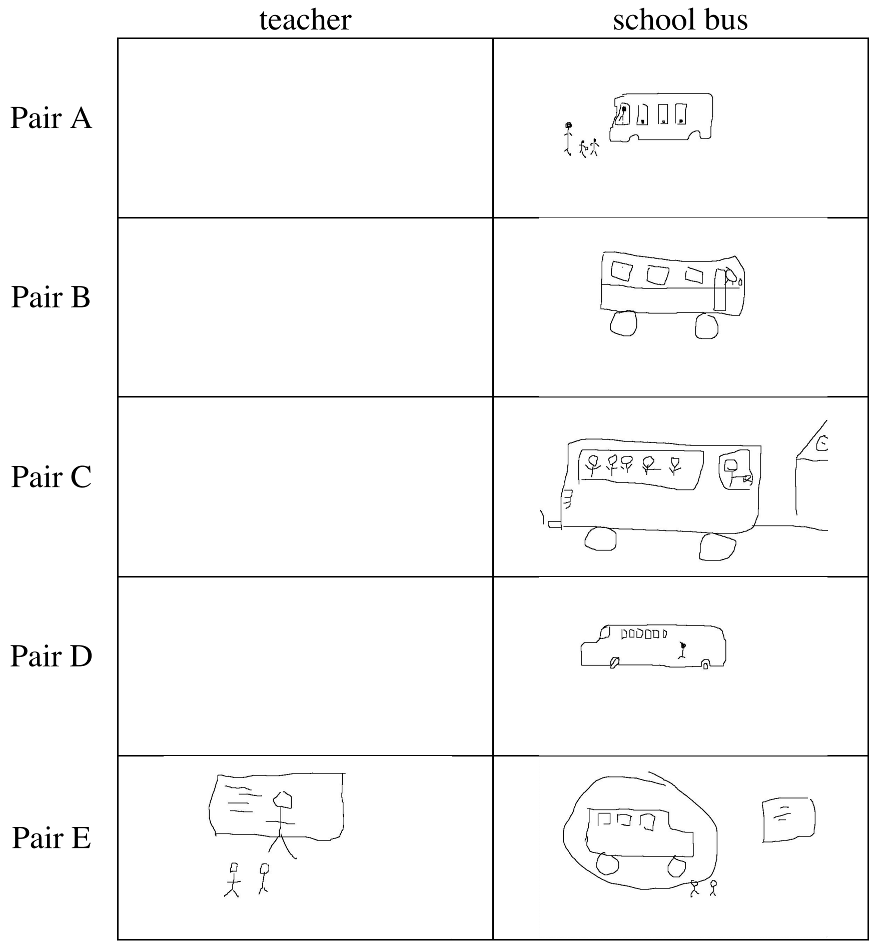


Figure 11. The first drawings of "school bus" from five different games. For pairs A – D, each was the first primary education item to be drawn in that game. In contrast, Pair E drew another primary education item ("teacher") before drawing "school bus". Their drawing for "school bus" appears to re-use elements from their drawing for "teacher": the chalkboard, in particular.

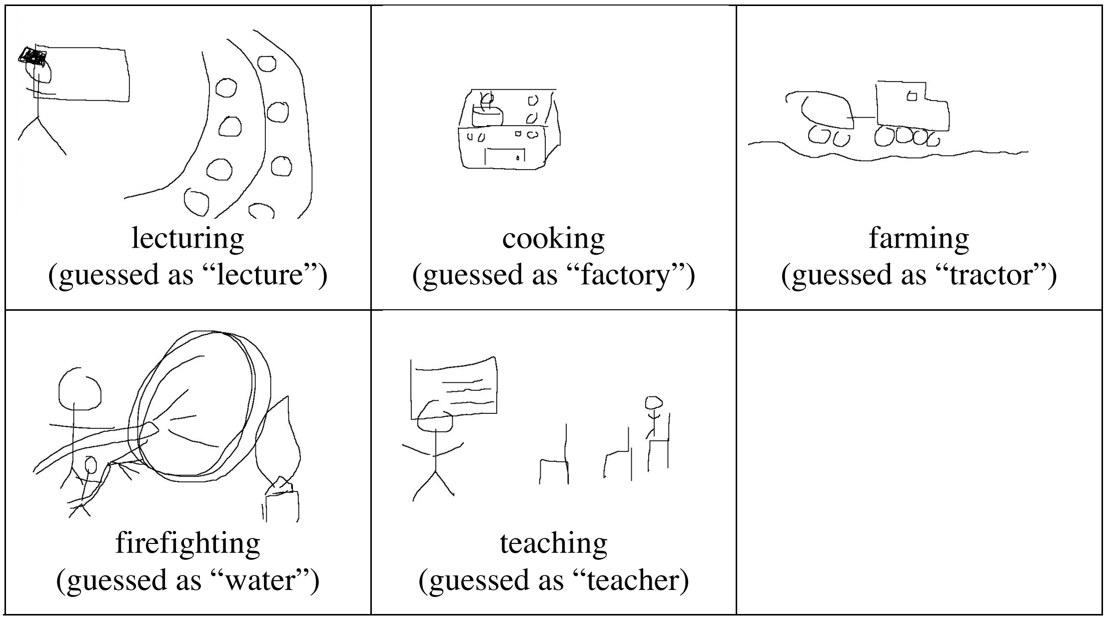
Note that most (at least 21 of the 26) initial drawings were produced after a related item (of either the same entity type or theme) had been drawn. Thus, interaction history can play a leading role in the earliest stages of the development of a sign system.

Increasing systematic compositionality due to interaction history. Another way we proposed that systematic compositionality could increase in a set of novel signs is that an element used in one sign could spread to others. There is a clear example of this in one of the games. Figure 12A shows the pair's drawings for the activities "teaching", "lecturing", "fire-fighting", "cooking", and "farming" in the beginning of their game. Note first that the pair is unsuccessful in each trial. Also note that the drawings have little in common with each other.

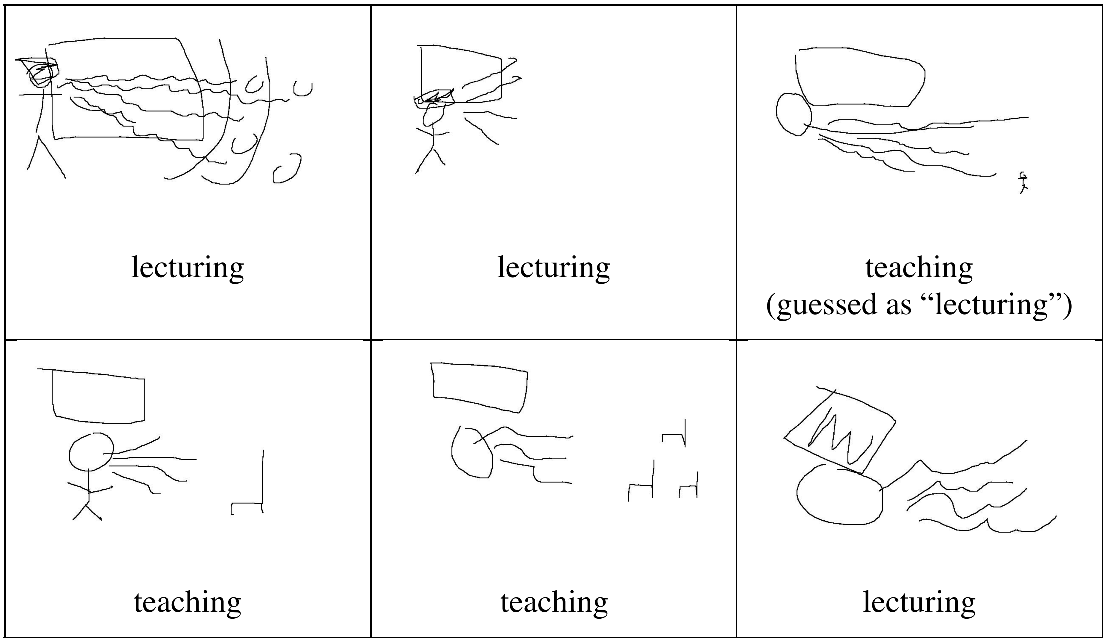
The next few of these activities to appear in their game happened to be "teaching" and "lecturing". Figure 12B shows these. The pair established a drawing for "lecturing" which includes squiggly lines. They then started to use those squiggly lines for "teaching" as well. At this point, two of their drawings for activities share an element (the squiggly lines).

In Figure 12C, we see this convention being extended to their drawings for "cooking", "farming", and "fire-fighting": each now contains the squiggly lines.

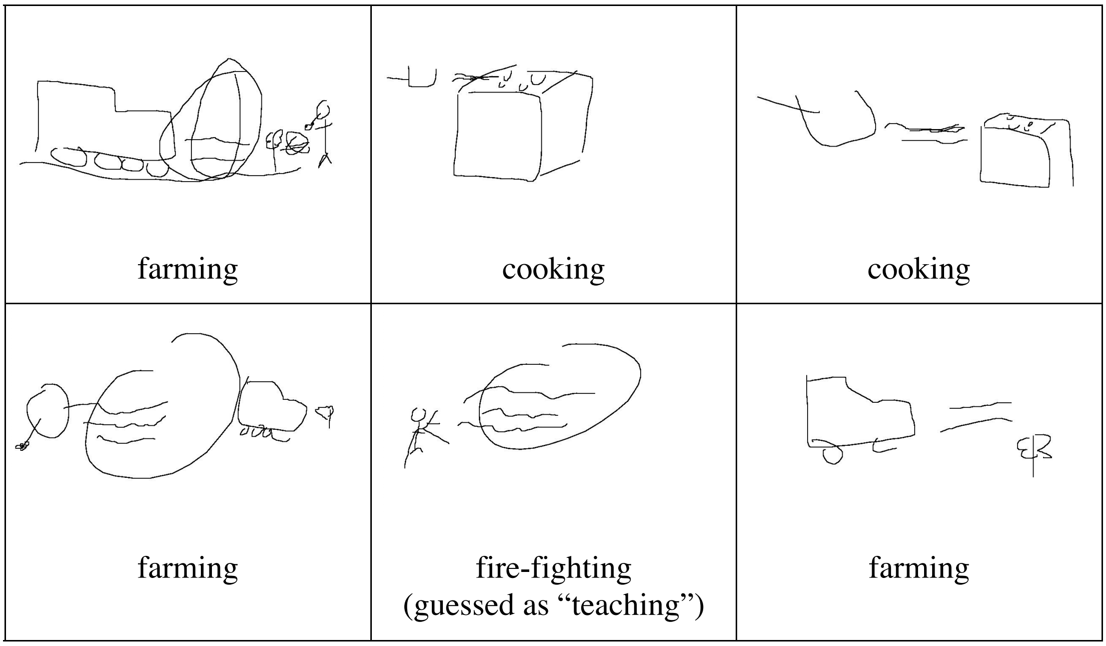
Figure 12D shows what was drawn for each of these items the last time it appeared in their game. Each of the drawings includes the squiggly lines.


lecturing
(guessed as "lecture")
cooking
(guessed as "factory")
farming
(guessed as "tractor")
firefighting
(guessed as "water")
teaching
(guessed as "teacher)

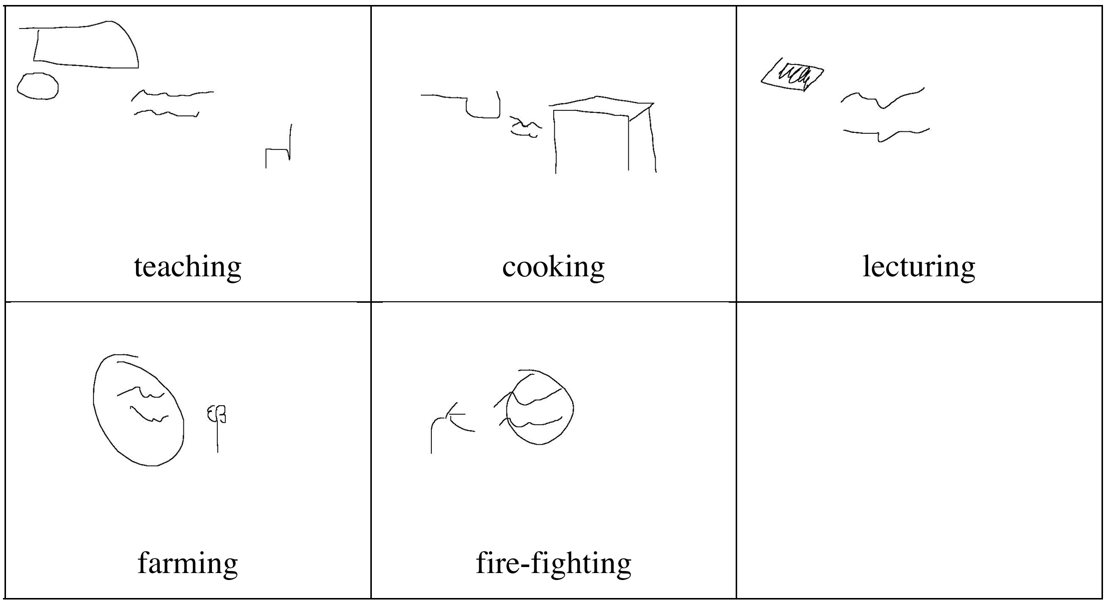
A


lecturing
lecturing
teaching
(guessed as "lecturing")
teaching
teaching
lecturing

B



C



D

Figure 12. One pair's drawings for the activities "teaching", "lecturing", "fire-fighting", "cooking", and "farming" in the beginning (A and B), middle (C), and end (D) of their game. They began drawing squiggly lines first for "lecturing". The squiggly lines then became established for "teaching" as well. Then, they extended the squiggly lines to their drawings for "fire-fighting", "cooking", and "farming". By the end of their game, the squiggly lines are a clear, distinct element shared by the drawings of these activities.

4.4.2 Systematic compositionality due to motivatedness

Another route to systematic compositionality that we discussed in Chapter 3 is categorization of motivated signs. Consider Figure 13, which shows how one pair drew four vehicles (“school bus”, “ambulance”, “tractor”, and “fire engine”) in the beginning and end of their game. In the beginning, there is no obvious shared element between the drawings – the drawings of the vehicles look different. For example, the vehicles are drawn to be of different lengths. By the end of their game, players have stopped making these distinctions, and this results in the drawings sharing an element.

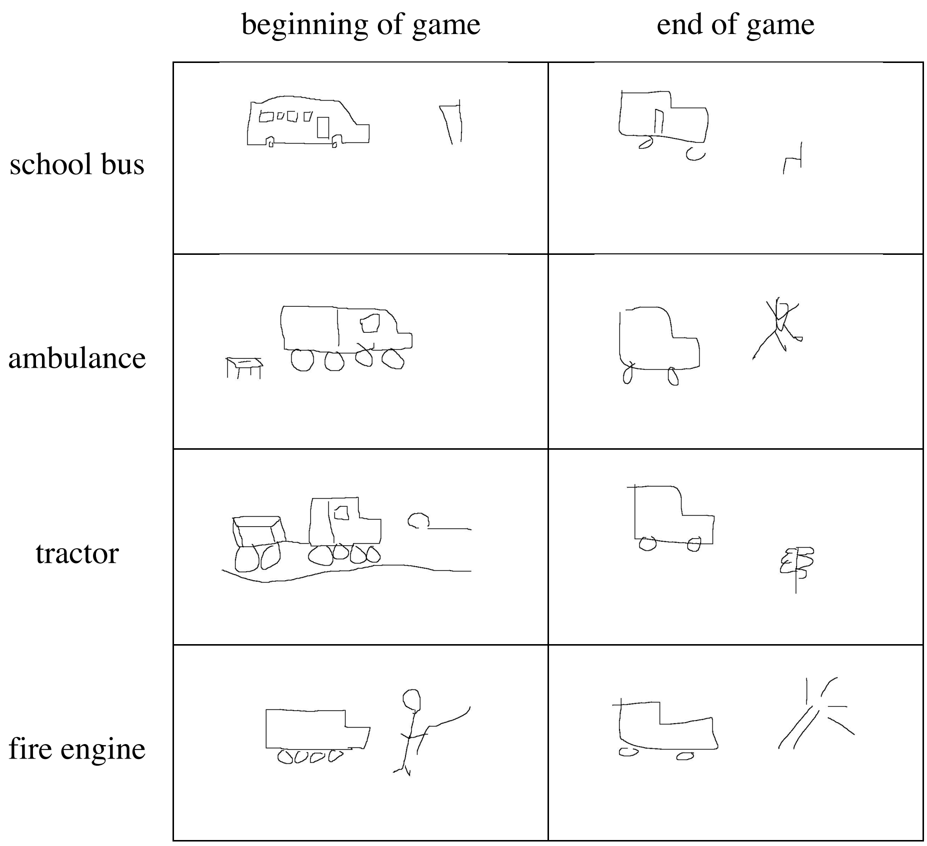


Figure 13. One pair’s signals for four different vehicles in the beginning and end of their game. Fine distinctions between the vehicles were represented in the drawings in the beginning of their game but collapsed by the end of their game.

4.4.3 Novel sets of signs were not completely systematically compositional.

The reader should not take from these demonstrations of systematic compositionality that systematic compositionality is inevitable in our communication game. The level of systematic compositionality in the sets never reaches 100%, and most sets include at least some completely idiosyncratic signs. Thus, participants did not appear to feel any artificial pressure to create systematically compositional sets of signs. For example, Figure 14 shows one pair's first drawings of items relating to farming, which have very little in common.

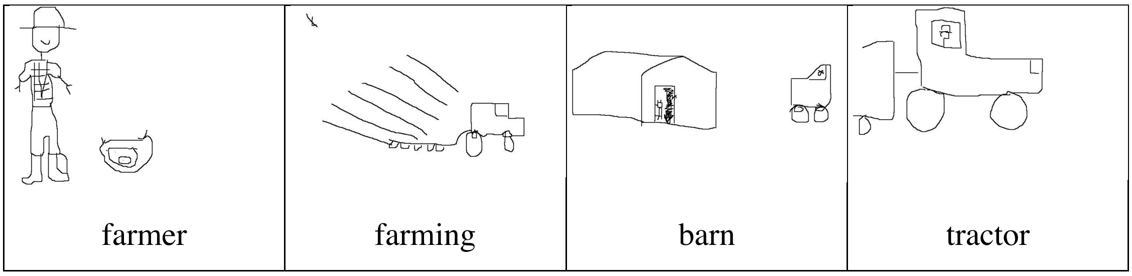
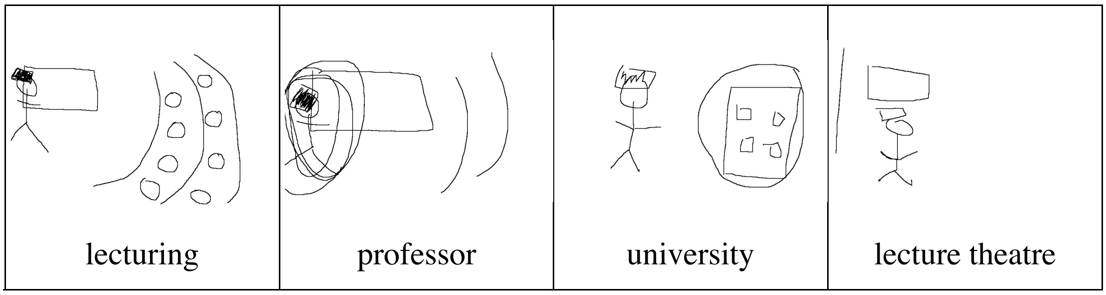


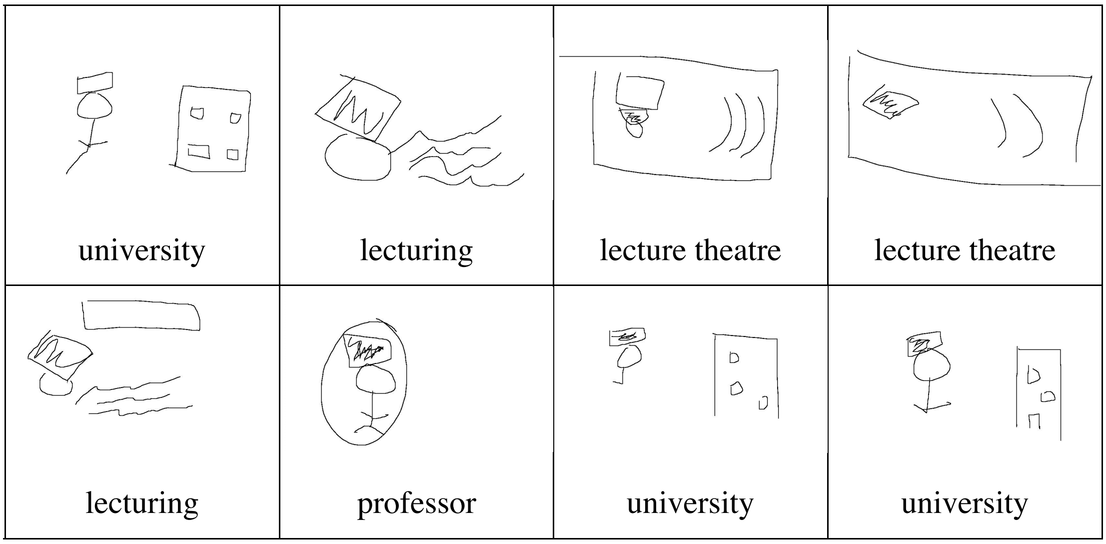
Figure 14. One pair's initial drawings for items relating to farming. They have little in common.

4.4.4 Sign *elements* can become arbitrary.

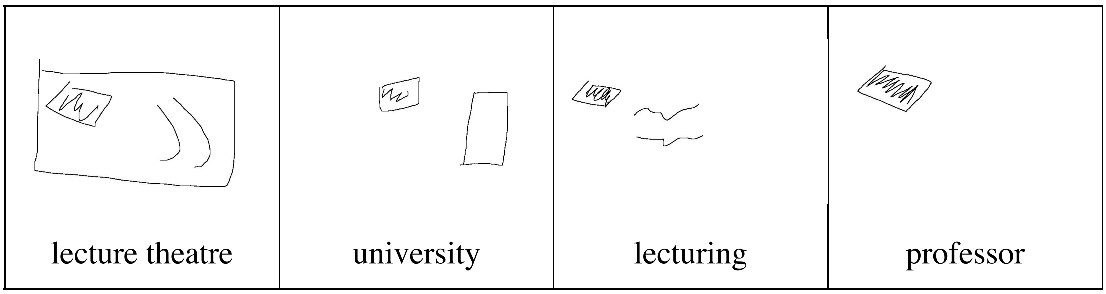
The other component to the parallel theory of the emergence of the systematic re-use of arbitrary elements, in addition to the idea that novel (motivated) signs can be systematically compositional, is the idea that systematic compositionality can be maintained as signs become arbitrary. (As a reminder, we aim to explain how the systematic re-use of *arbitrary* elements arises, not just how the systematic re-use of any elements arises.) We did find this in Experiment 1. Consider Figure 15, which shows some of one pair's drawings for university items over the course of their game. Figure 15A shows the pair's initial drawings for university items. The set is highly systematically compositional, in that each contains a stick figure with a mortarboard on its head, which (in three of the four drawings) is standing next to a chalkboard. Figure 15B shows the drawings for the next eight university items to appear in the game (after the last initial drawing). This shared element is becoming



A



B



C

Figure 15. One pair's initial drawings of university items (A), drawings of university items produced in the middle of the game (B), and final drawings of university items (C). The initial drawings for university items are highly systematically compositional, in that they share an element. Over the course of the game, this element becomes arbitrary.

arbitrary – the body, head, and chalkboard are being eliminated. Figure 15C shows the pair's final drawing of each university item. The shared element is now very arbitrary – a filled-in diamond without the context of a body, head, and chalkboard has a weaker connection to the concept of university education. Crucially, the initial systematic compositionality of the signs was retained while the signs became arbitrary.

4.4.5 Sometimes, whole signs become arbitrary.

Lest the reader take from this that systematic compositionality was always maintained as signs became arbitrary, we present an example of whole signs – not sign elements – becoming arbitrary in Figure 16. Figure 16A shows this pair's initial drawings for farming items. They are systematically compositional, in that a barn and tractor recur across the drawings. Figure 16B shows the drawings for the next eight farming items to appear I the game (after the last initial drawing). These drawings for farming items share elements with each other to a lesser extent. Figure 16C shows the pair's final drawings for farming items, which exhibit little systematic compositionality. Notice in particular their idiosyncratic drawings for "barn" and "tractor" – they don't share any elements at all with the other drawings.

4.5 Summary

We've presented an experiment in which the systematic re-use of arbitrary elements emerges. Final sets of signs are systematic, and becoming more arbitrary.

The serial theory of the emergence of the systematic re-use of arbitrary elements predicts that a novel set of signs will first become arbitrary and then develop systematic compositionality. The results from Experiment 1 support another possibility: the parallel theory of the systematic re-use of arbitrary elements. Novel signs are systematically compositional immediately, and this systematic compositionality is maintained as the signs become arbitrary.

Examples from Experiment 1 provided evidence for the various routes to the systematic re-use of arbitrary elements that we proposed in Chapter 3. In particular, the shared interaction history between players appears to play an important role in the emergence of systematic compositionality in their signs. However, we also showed that our results are not trivial – subsets of novel signs sometimes exhibit very little systematic compositionality, and subsets of signs sometimes do become less systematically compositional as they become arbitrary.

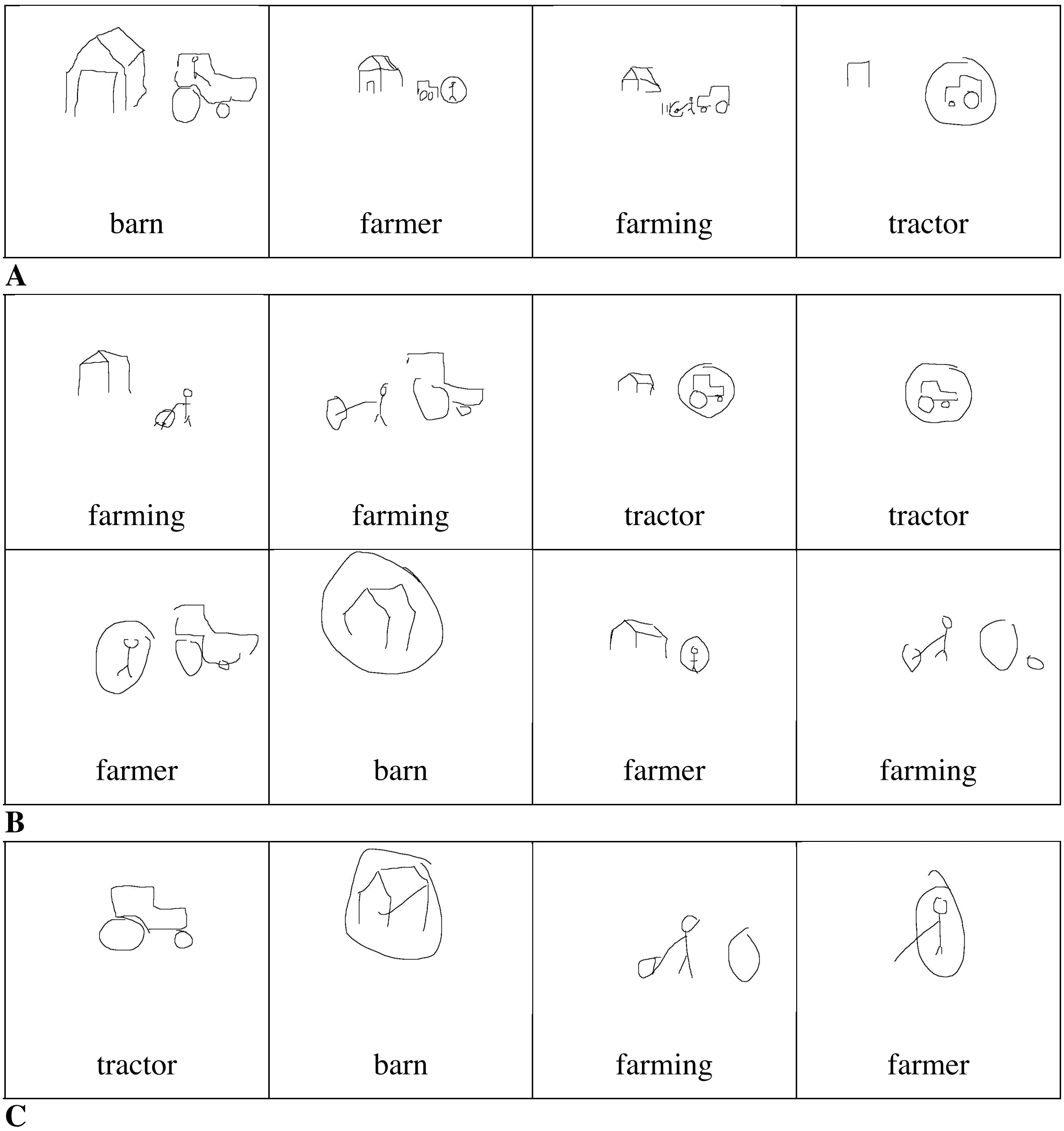


Figure 16. One pair's initial drawings of farming items (A), drawings of farming items produced in the middle of their game (B), and final drawings of farming items (C). The drawings for farming items are systematically compositional initially, but become less so over the course of the game.

Chapter 5

Experiment 2: Generalization

The previous chapter presented Experiment 1, which demonstrated how the systematic re-use of arbitrary elements arises in novel sets of signs created by interacting communication partners. In this chapter, we begin to explore how such sign systems could be extended beyond individual pairs.

5.1 Rationale

One might argue that the sets of signs that emerged in dyads in Experiment 1 could not lead to proper language, which is shared by large communities of people and also persists through generations. Experiment 2 was designed to test whether new participants – who did not play any role in (indeed, did not even observe) the creation of the sets of signs – can nonetheless detect the systematic compositionality in them and generalize from it.

5.2 Method

In a nutshell, new participants learn subsets of systematically compositional (or control) sets of signs, and are asked to draw what they think the original participants drew for the missing items. We then test whether two participants who learned the same set of signs draw the missing items the same way. Specifically, we test whether their two sets of generalizations are more similar to each other than either is to sets of generalizations made by participants who learned *different* sets of signs (i.e. sets of signs produced by different original pairs). If so, we reason that there's information in each observed subset of signs about the missing signs. In particular, the observed subset is systematically compositional, in that signals for similar meanings share elements, and so a new participant can guess that an original pair's signal for a similar item also included that element. We will also know that the information is specific to the set of signs – it's not that new participants all make the same guesses about how to draw the missing items. In sum, this generalization experiment will establish whether new participants can detect and generalize from the (set-specific) systematic compositionality in others' signs.

5.2.1 Participants

24 University of Edinburgh students participated in exchange for £12. All were native British English speakers. None had a background in Linguistics.

5.2.2 Stimuli

There were 12 stimulus sets: two different subsets of six sets of signs (four Normal sets, a Mixed set, and a Random set).

The Subsets. Each subset contained drawings for 21 of the 26 core items. We call these 21 the Seen signs because these are the signs the generalization participants learned. The five missing signs we call Unseen; the generalization participants were tested on these. The five Unseen items were chosen at random, with the restriction that each entity type (person, building, etc.) and theme (primary education, fire-fighting, etc.) included at most one Unseen item. The Unseen items for subset #1 were: "chef", "hospital", "fire-fighting", "school bus", and "lecture theatre". Those for subset #2 were: "farmer", "hospital", "lecturing", "fire engine", and "classroom".

Normal. Four of the sets of signs were Normal. These were the two most systematically compositional Final sets of signs from Experiment 1 and the two most systematically compositional Final sets of signs from Experiment 3. We present Experiment 3 in the next chapter. For now, it suffices to note that participants in Experiment 3 play the same game as in Experiment 1 but spend 15 minutes before the game familiarizing themselves with another pairs' signs. To review, a Final set of signs consists of what was drawn for each core item the last time that item appeared in a pair's game. As the Normal sets we chose were highly systematically compositional, the drawings for similar items will often share elements. For example, Figure 17 shows subset #1 of Normal set #2. In it, we see that each drawing for an item relating to university education (in the third row) contains diagonal lines in the top left corner and a stick figure to the right of them. Similarly, each drawing for a room (in the fifth column) includes a large box around a scene.

Again, we predicted that one participant's generalizations would be more similar to the generalizations made by the participant who learned the *same* set of signs than they are to generalizations made by any participant who learned a *different* set of signs. We foresaw two possible objections to interpreting this result as indicating that the new participants are detecting and using systematic compositionality in the original pairs' signs.

First, participants might base their guess for an Unseen item on one particular Seen item, rather than on all Seen items similar in meaning. For example, when guessing what the original participants drew for "farmer", we assume that the generalization participants look at all the drawings for items that relate to farming and all the drawings for activities/situations. However, it's possible that both generalization participants look only at the drawing for "farming" and modify that. If so, the similarity of their generalizations for "farmer" would not imply systematic compositionality (a re-use of drawing elements across drawings for similar items) in the original pair's set of signs, for they could make similar generalizations even if the set of signs exhibited no systematic compositionality.

Second, participants might base their guess for an Unseen item on *all* of the Seen items, not just those that are similar in meaning to it. For example, if the original pairs had particular drawing styles (e.g., a pair tended to use more or less ink, drew generally round vs. generally straight lines, tended to be sloppy and quick vs. careful), and the generalization participants used this information, then generalizations from participants who learned signs from the same original pair

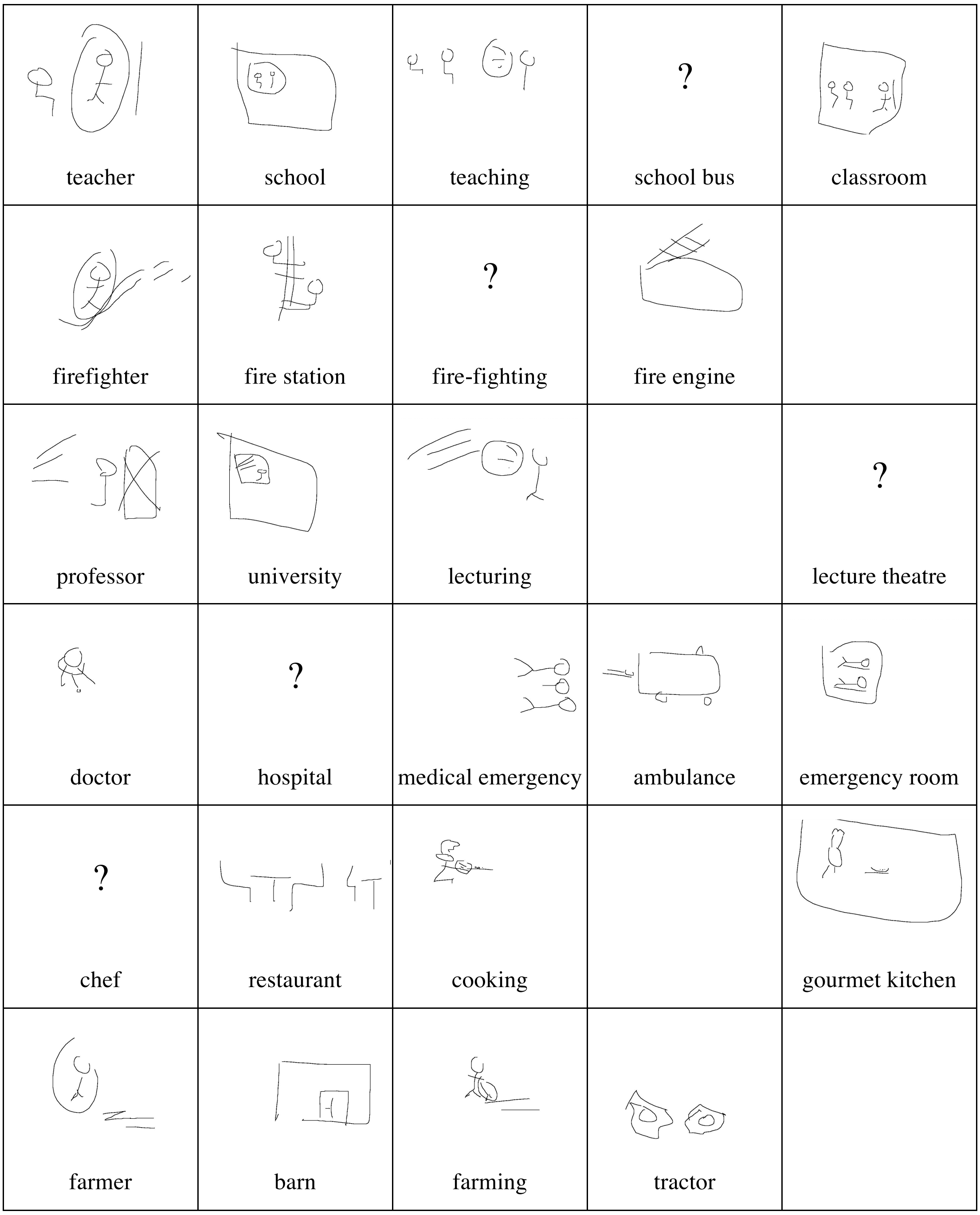


Figure 17. Subset #1 of Normal set #2. Each image shows what was drawn the last time that item appeared in this pair's game. Two generalization participants learned the signs shown and then were asked to guess what the original participants drew for the five Unseen items. The Normal sets are highly systematically compositional. For example, in this set, drawings for items relating to university education include three diagonal lines and a stick figure and drawings for rooms include a large box around a scene, so a generalization participant can make an informed guess of how the original participants drew "lecture theatre".

would be more similar to each other than they would be to generalizations from other participants.

We constructed a Mixed set to address the first possible objection and a Random set to address the second.

Mixed. We constructed a Mixed set, which exhibited virtually no systematic compositionality. It included signs from all four Normal sets, and so was analogous to the Mixed Final sets in Experiment 1, which exhibited little systematic compositionality. However, this Mixed set was composed of signs from only four sets of signs. Since there are more than four entity types and more than four themes, we additionally ensured little systematic compositionality by choosing signs so that a particular signal element appeared no more than once per row or column. For example, no two drawings of items relating to university education included diagonal lines in the top left corner. We also chose signs that were clear and understandable (rather than messy or obscure), so that participants could copy elements in them. Figure 18 shows subset #1 of the Mixed set.

Generalizations made by participants who learn this Mixed set will be similar to the extent that participants are basing each guess for an Unseen item on one particular Seen item.

Random. We constructed a Random set, which exhibited little systematic compositionality but preserved information about how an original pair drew items in general. The drawings in the Random set were exactly those of one of the Normal sets (Normal set #4), mapped at random to the items. For example, the original pair's drawing for "gourmet kitchen" was said to be for "firefighter". Figure 19 shows subset #1 of the Random set.

If generalizations from participants who learn the same Normal set are more similar than generalizations from participants who learn the Random set, we know that the

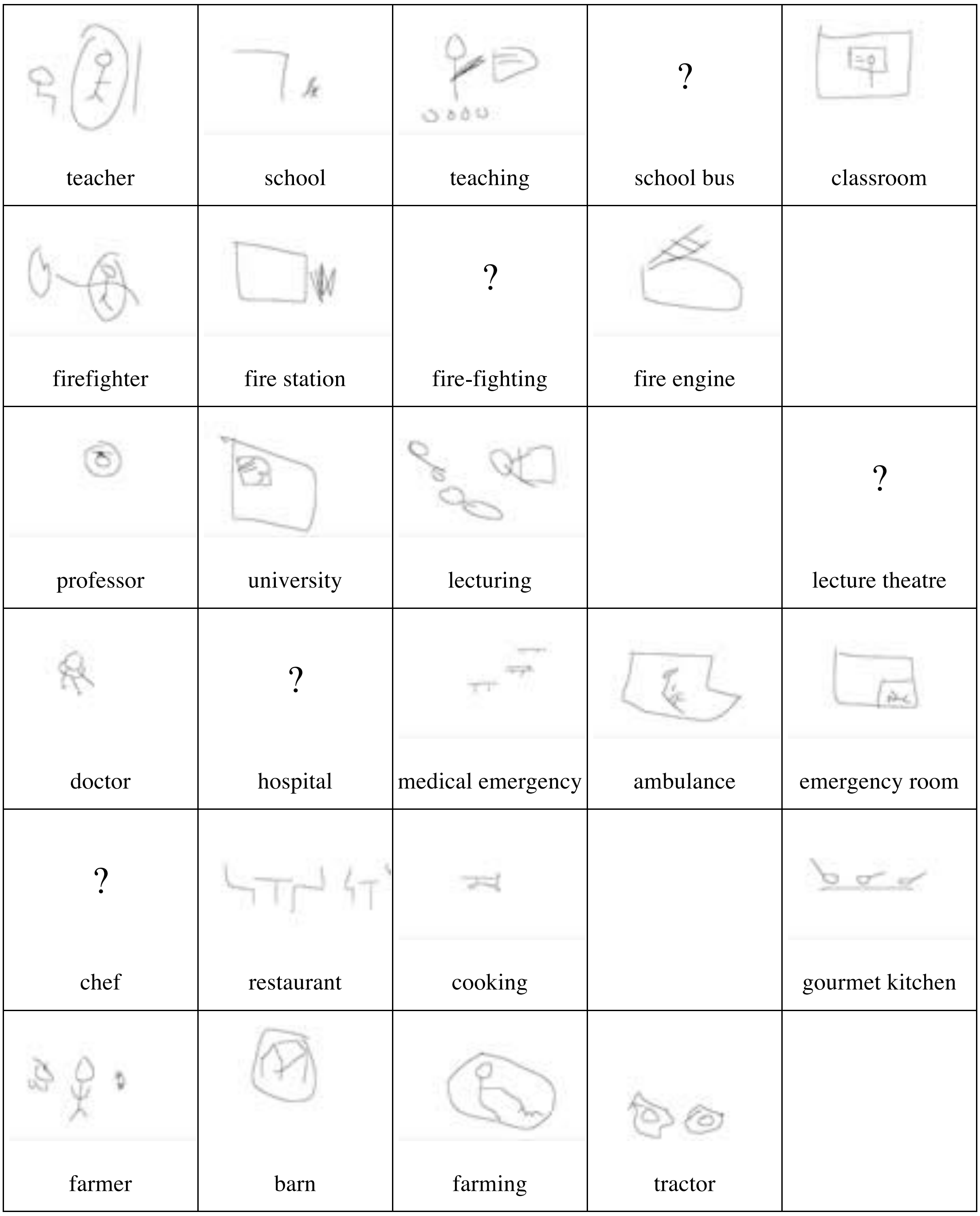


Figure 18. Subset #1 of the Mixed set. Each image shows what was drawn the last time the item appeared in an original pair's game, but the images are drawn from *different* games. Any similarity in generalizations made by participants who learn this set is due not to systematic compositionality but perhaps to the participants generalizing from the same (single) Seen item.

Figure 19. Subset #1 of the Random set, which consists of the drawings from one Normal set mapped randomly to the items. Any similarity in generalizations made by the participants who learn this set is not due to systematic compositionality, but rather to the original pair having a distinctive drawing style.

latter participants are using information about the mappings between drawings and items, not just information about the drawings.

Thus there were 12 stimulus sets in total: two subsets each of six sets of signs (four Normal, one Mixed, and one Random). Each participant learned one, assigned to him at random. Each set was learned by two participants.

5.2.3 Procedure

The drawings were presented to the participant on a computer monitor, as a folder of images labelled with the items. The images appeared in the folder in random order.

Each subject had 15 minutes to learn the 21 drawings in any way he wanted. He was told he would be tested: afterwards, the experimenter would present him with some items and he would have to draw what the original participants drew for them. Microsoft Paint, a simple graphics painting program, was running so that he could practice drawing the items if he wanted.

Following this training, the first test was given. The participant was asked to draw the five Unseen items in turn, presented to him in random order, using Microsoft Paint. The participant was told that these were items that he did not learn, and that his task was to guess (by drawing) what the original participants drew for them. The participant was encouraged to take as much time as he needed, was allowed to look at all of the Seen items while drawing the Unseen items, and was encouraged to explain to the experimenter what he was guessing and why during the test. This first test was a practice test for the second test.

For the second test, the participant was shown the Seen drawings arranged in a table on a printout. The experimenter explained that the columns were entity types (people, buildings, etc.) and the rows were themes (primary education, fire, etc.) The experimenter told the participant that looking at the drawings this way might give him more information. The procedure for the second test was otherwise identical to that for the first test. (To be clear, the five Unseen items were in a *different* random order in the second test than they were in the first.)

The full instructions to the participants are included as Appendix C.

5.3 Results

5.3.1 Example generalizations

Figure 20 shows the sets of generalizations made by six of the participants. Two of these participants learned one of the Normal sets (Normal set #2), two learned the Mixed set, and two learned the Random set. Each of the six participants learned drawings for the same items (subset #1 of their respective sets of signs) and was tested on the same Unseen items. Yet only the two who learned the (same) Normal set appear to be drawing the items the same way. The participants who learned the Mixed or Random sets do not appear to agree on how the original participants drew the Unseen items.

5.3.2 Similarity judgments

Six volunteers provided similarity judgments on the sets of generalizations produced in the second test of the experiment. For each of the two subsets, there were 12 sets of generalizations, one per participant:

Participant A learned subset #1 of Normal set #1
Participant B learned subset #1 of Normal set #1
Participant C learned subset #1 of Normal set #2
Participant D learned subset #1 of Normal set #2
Participant E learned subset #1 of Normal set #3
Participant F learned subset #1 of Normal set #3
Participant G learned subset #1 of Normal set #4
Participant H learned subset #1 of Normal set #4
Participant I learned subset #1 of the Mixed set
Participant J learned subset #1 of the Mixed set
Participant K learned subset #1 of the Random set
Participant L learned subset #1 of the Random set

For each volunteer, the 12 sets of generalizations (of five drawings each) for a given subset were printed onto strips of paper, and these strips of paper were arranged on a table in random order. First, the volunteer paired the sets by similarity, i.e. he formed most-similar pairs of sets. He then rated the degree of similarity between the sets that he paired, with 1 meaning there was no similarity at all between the drawings, and 10 meaning the drawings were exactly the same. This procedure was repeated for the 12 sets of generalizations made from the second subset.

The full instructions to the similarity judges are included as Appendix D.

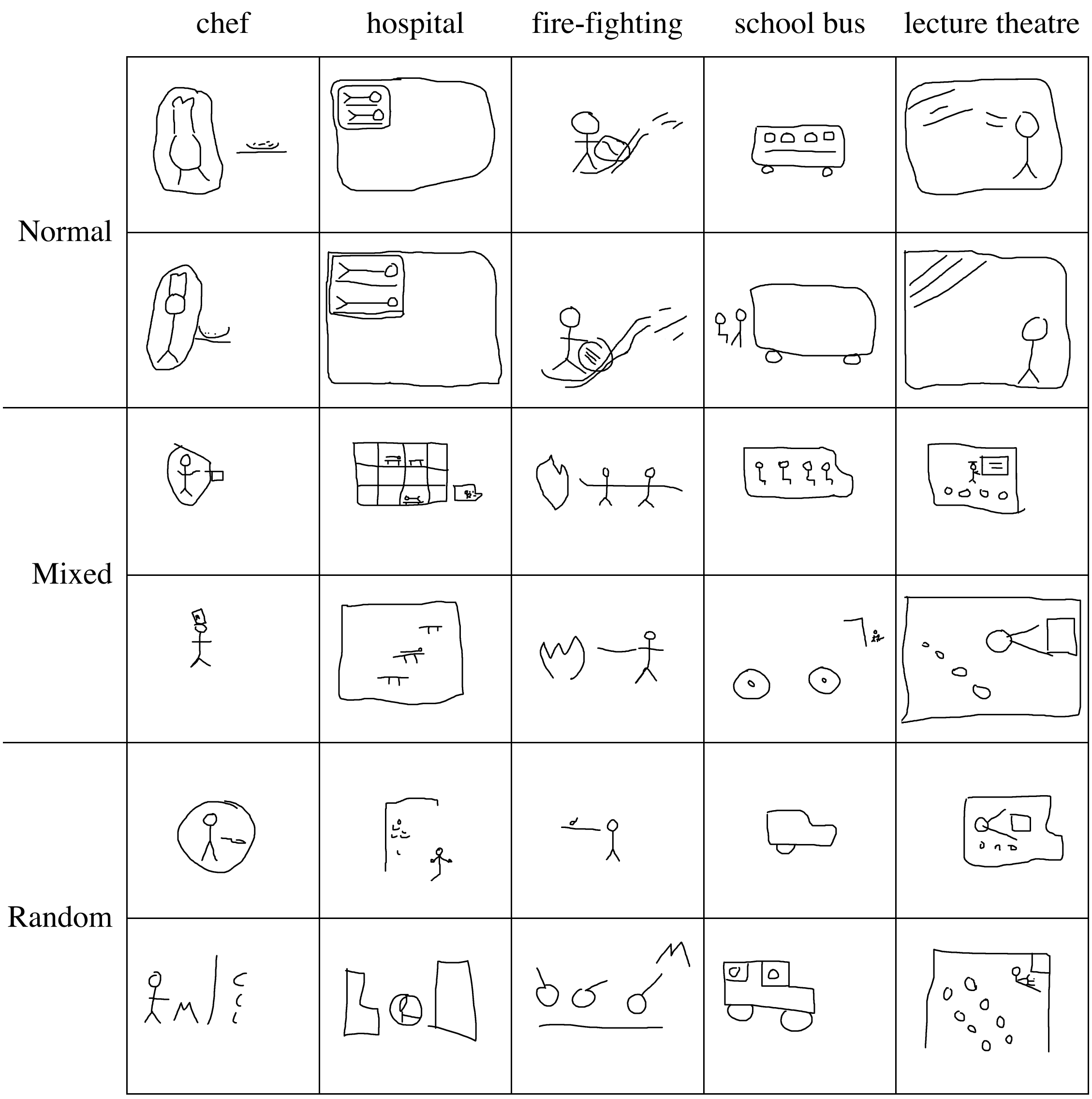


Figure 20. Sets of generalizations from six of the participants: two who learned the same Normal set, two who learned the Mixed set, and two who learned the Random set. Each learned signs for the same items and was tested on the same Unseen items. Despite this, only the generalizations from the two who learned the (same) Normal set are very similar.

5.3.3 Results

We call a pairing of the generalizations produced by two participants who learned the same set of drawings "correct". Table 1 shows, for each subset of each set, how many of the six judges made the correct pairing. No judge ever paired generalizations made by the two participants who learned the Mixed set or generalizations made by the two participants who learned the Random set as most similar. In contrast, generalizations produced by two participants who learned the same Normal set were often paired as most similar. For example, all six judges paired the sets of generalizations made by the two participants who learned subset #1 of Normal set #1 as most similar. A Chi-square test confirmed that the frequency of correct pairings differed between the three types of set (Normal, Mixed, and Random): χ^2 (2, N = 72) = 36, p < 0.001.

	Normal					
	#1	#2	#3	#4	Mixed	Random
Subset #1	6	6	3	4	0	0
Subset #2	5	2	5	5	0	0
Total	11	8	8	9	0	0

Table 1. Correct pairings of generalizations. Each cell displays the number of judges (out of six) who paired the correct pairing as most similar. No judge paired the generalizations from two participants who learned the Mixed set or the Random set as most similar, while many paired generalizations produced by two participants who learned the same Normal set as most similar.

Generalizations from participants who learned the same Normal set are more similar than generalizations from the participants who learned the Mixed set and generalizations from the participants who learned the Random sets. But are the generalizations from participants who learned the same Normal set more similar than generalizations incorrectly paired? While the task of the similarity judges is to pair the generalizations according to similarity, they must pair all of the generalizations. This means they may be forced to pair generalizations they do not find particularly similar. We elicited the similarity ratings from the judges to test that correct pairings (of Normal sets, as correct pairings were never made of the Mixed and Random sets) are indeed more similar than incorrect pairings. Table 2 shows the mean similarity ratings for each type of pairing. Correct pairings of generalizations (from the same Normal sets) are rated more similar than other (incorrect) pairings. ($M_{Correct} = 7.42$, $SD = 1.48$; $M_{Incorrect} = 4.42$, $SD = 2.13$) A Mann–Whitney U Test[19] confirmed this difference between the similarity ratings for correct and incorrect pairings ($p < 0.001$).

Correct		Incorrect		
both Normal	both Control	both Normal	both Control	one Normal, one Control
7.42	(none)	4.33	4.00	4.06

Table 2. Mean similarity ratings for correct and incorrect pairings of generalizations. Judges rated their correct pairings of generalizations produced by participants who learned Normal sets as most similar.

Thus, generalizations produced by participants who learned the same Normal set were paired as most similar often, while generalizations made by participants who learned either the Mixed or Random set were never paired as most similar, by any of the six judges, for either of the two subsets. Further, judges rated the correct pairings of generalizations produced by two participants who learned the same Normal set as more similar than any incorrect pairings.

[19] The similarity ratings for the incorrect pairings failed standard tests for normality.

This tells us:

- There's something in each of the Normal sets of signs, but not in the Mixed or Random sets of signs, that gives the participants information on how to draw the Unseen items.

- This information is specific to the set, for a participant's generalizations are judged to be most similar to the generalizations from the participant who learned the *same* Normal set, not just any Normal set.

- Participants are not basing their guess for a particular item (say, "lecture theatre") on one particular Seen item (say, "university"). If they were, generalizations from the two participants who learned the Mixed set would be paired as most similar, as would the generalizations from the two participants who learned the Random set[20].

- Generalizations from participants who learned the same Normal set are not most similar simply because they are detecting drawing styles unique to the original pairs. If this were the case, the generalizations from participants who learned the Random set would also have been paired as most similar, for the Random set consists of signs produced by the same original pair.

5.4 Observations

Here we make some observations about how the systematic re-use of arbitrary elements produced by a pair was adopted by others.

---

[20] It's possible that each participant bases a guess on one Seen item, but that that Seen item varies across participants. The difference between Normal sets on the one hand and Mixed and Random sets on the other still indicates that the Normal sets were highly systematically compositional – the two Seen items the Normal participants based their guesses on must have something in common.

5.4.1 Systematic compositionality

A qualitative analysis of the data generated in Experiment 2 provides further evidence of the systematic compositionality in the original pairs' signs.

Choice of subset may not affect whether participants reproduce the sign elements. One might wonder whether the choice of subset of a given set affected how participants generalized from it. The more systematically compositional a set of signs, the less the choice of subset should affect generalizations made from it. For example, in Normal set #2, each drawing of an item relating to university education includes three diagonal lines in the top left corner. A new participant could learn any subset of these and still detect the systematic compositionality.

There is some evidence that the same sign elements were detected, regardless of subset. For example, the four Normal sets of signs each signalled that an item relates to university education using a different element: a head with mortarboard, three stripes for lecture theatre seats, a diagonal row of small circles, and a man with a bowtie. Figure 21 shows how the participants who learned the two different subsets of the four Normal sets drew their item relating to university education. Those who learned subset #1 were asked to guess what the original participants drew for "lecture theatre", while those who learned subset #2 were asked to guess what the original participants drew for "lecturing." The generalizations of "lecture theatre" share an element (such as a head with a mortarboard) with the generalizations of "lecturing" when those participants learned signs from the same original pair. Regardless of the subset observed, participants who learned Normal set #1 include a head with a mortarboard in their drawings, the participants who learned Normal set #2 include three diagonal lines in the top left corner, the participants who learned Normal set #3 include a diagonal row of small circles, and the participants who learned Normal set #4 include a man with a bowtie. This is further evidence of the strength of the systematic compositionality in the original sets of signs.

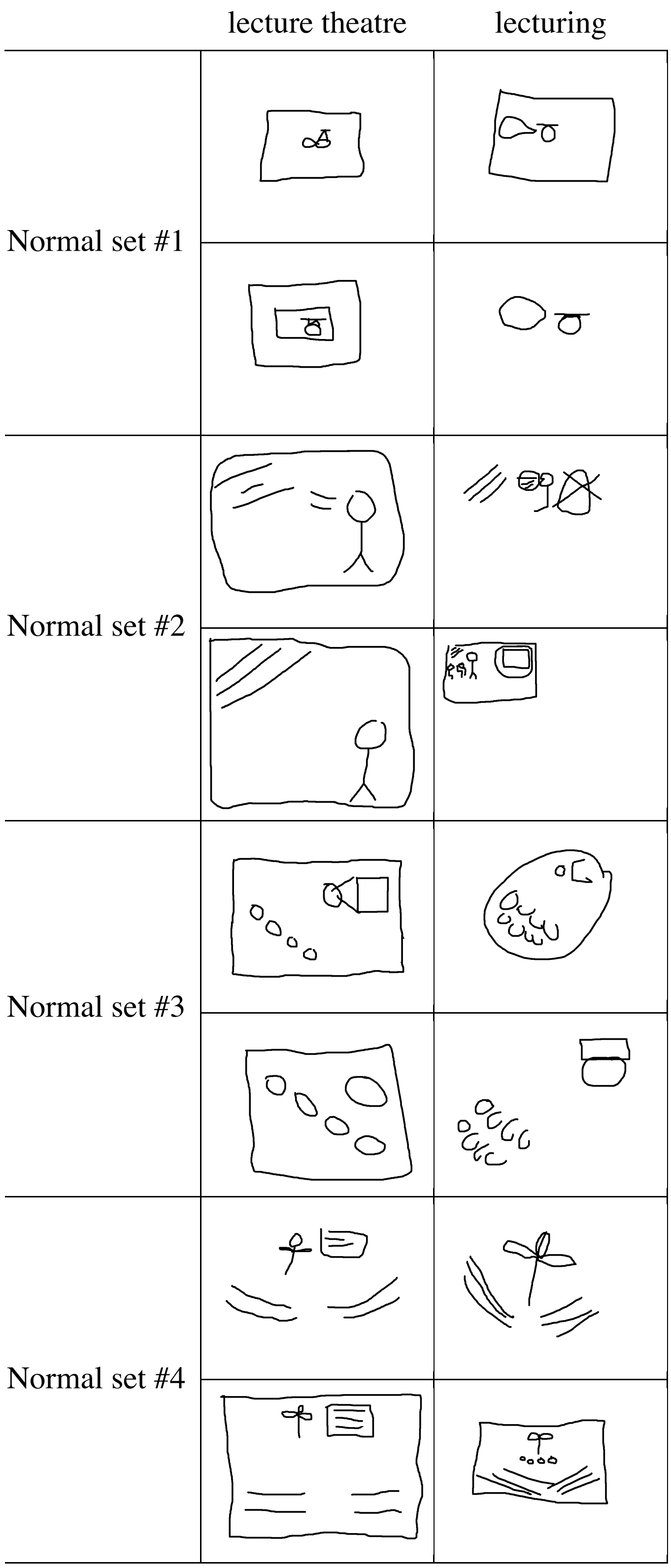


Figure 21. Generalizations for two different items relating to university education. Despite learning different subsets thereof, participants who learn the same set of signs include the same elements in their generalizations.

The experiment was not designed to allow a test of this, though. Recall that each subset's five Unseen items were chosen at random, in order to eliminate any experimenter bias. They were:

- subset #1: chef, hospital, fire-fighting, school bus, lecture theatre
- subset #2: farmer, hospital, lecturing, fire engine, classroom

The first problem is that the two Unseen sets share an item: hospital. Thus we don't know whether participants who learn the drawings for different subsets of medical items and are tested on different new medical items would nonetheless reproduce the same drawing element. The same holds for a drawing element indicating that an item is a building.

The second problem is that only one of the Unseen sets contains an item that relates to cooking, and only one of the Unseen sets contains an item that relates to farming. Obviously, we cannot test whether the choice of subset affects how people generalize to new cooking or farming items in this case.

Participants' verbal feedback indicates that they are detecting systematicity. Another piece of qualitative evidence for the strength of the systematic compositionality in the sets of signs comes from participants' comments while they're making their generalizations. Despite never being explicitly instructed to look for systematicity in the set of signs they're learning, they very often explicitly state what rules they're following. Examples include:

- [The original participants have] "obviously got some kind of system going."
- [The few lines in a circle are] "code for doing."
- "Lines coming out of him means doing things."
- "People have quite often got a circle around them, but not when it has to do with "ing"."
- "When it's the room that's the word to be conveyed, put a box around the entire picture."
- "…put the other specific on the left."
- "The difference between verb and noun seems to be what gets circled."

- "The guy with the bump on his leg features as indicating some kind of illness."
- "That's their vehicle."
- "...their school and university principle"
- [The] "board symbolizes it's something educational."

5.4.2 Arbitrariness

The relative arbitrariness of the signs affected whether the new participants used the systematic compositionality in them. Their comments indicate that they are sometimes willing to copy somewhat arbitrary elements but other times not.

Participants sometimes copy things that don't make sense to them. For example, one participant drew a farmer "with four arms" even though he admitted not knowing *why* he would have four arms, even joking that the farmer might live in Chrnobyl. (The participant was using the drawing of "farming" that he learned, in which the original participants put emphasis on the arms to indicate an activity.) Another participant reported that the original pair's drawing for "professor" looks like a moth, and another that their doctor has "circle hands", but they copied these elements nonetheless.

Participants sometimes change the drawing elements they learned to make more sense to them. For example, one participant used an arrow "to highlight [that it's a] room because circling it would be strange" – despite the fact that she knew that the original participants never used an arrow and often used circles. Another said all the original participants' vehicles looked a certain way but drew school bus bigger than that because it "needs a bit of space for people". Another said she copied the drawing of farming she learned except for a circle off to the side which never made sense to her. Participants may assume they are misunderstanding the original pair's drawings somehow.

Participants sometimes report producing motivated drawings for the items. For example, one participant said his drawing of "hospital" was "definitely going to have a big building because a hospital *is* a big building" so that's the first thing that came to his mind. For "lecturing", one participant put a professor in the drawing because "*someone* has to do the lecturing." Even though three out of four medical items he learned had a patient with an injured knee, one participant put a doctor by his hospital instead of the patient with injured knee because the "natural assumption would be that the doctor would be in the picture as well."

Participants sometimes misunderstand drawing components. For example, the second participant to have learned the drawing of "farming" that included the man with heavy emphasis on his arms understood that man to be crops. When he drew "farmer", he drew a new stick figure next to these crops.

Thus, generalizing from a set of systematically compositional signs that have become somewhat arbitrary is no trivial task.

5.5 Summary

This thesis aims to address how the systematic re-use of arbitrary elements can arise in a language. We saw it arise in Experiment 1 within dyads. Experiment 2 has shown that this could be the start to proper language, which is used by entire communities of people and over generations. While the systematic compositionality in the sets of signs may have emerged within a dyad, people who did not help create the signs – or even observe their creation – nonetheless can detect it and use it. Specifically, new participants are able to detect the systematic compositionality in the original sets of drawings and reliably generalize from it.

Chapter 6

Experiment 3: Observation

This thesis considers the evolution of arbitrariness and the evolution of systematic compositionality together, and proposes that the two are not serial developments but rather parallel. In Chapter 4 we presented Experiment 1, which showed the emergence of the systematic re-use of arbitrary elements in novel signs created by pairs of interacting communication partners. The emergence of systematicity and the emergence of arbitrariness were parallel, not serial, developments: systematicity emerged immediately, while signs were still highly motivated. One might argue that the communication systems that emerged in Experiment 1were just conceptual pacts that could not extend beyond the dyad, i.e. that these communication systems are not relevant to proper language, which is used by large communities and generations of people. So in the previous chapter, we began to explore whether these communication systems could be extended beyond the dyad. We presented Experiment 2, the generalization experiment, that showed that people who had no hand in – indeed, did not even observe – the development of the sets of signs can nonetheless detect the systematic compositionality in them and generalize from it when producing signs for unobserved items. In short, people *can* adopt the systematic reuse of arbitrary elements that emerged from motivated signs.

In this chapter, we present an experiment which tests whether people actually *do* adopt the systematic reuse of arbitrary elements that emerged, when faced with the task of not learning and reproducing, but rather communicating.

6.1 Rationale

The goal of the experiment was to test whether the systematicity that developed in Experiment 1 – that which persisted from motivated signs –persists through generations of participants whose task was communication.

6.2 Method

Experiment 3 is almost identical to Experiment 1. The first exception is that the participants spend 15 minutes observing another pair's signs before playing their own game. The second exception is that the pairs of participants are organized into chains of generations, so that a pair observes signs produced by a pair who played before them, who in turn observed signs from a pair who played before them, and so on.

Below, we note only the differences from Experiment 1.

6.2.1 Participants

24 University of Edinburgh students were recruited.

Participant pairs were organized into 4 chains of 4 generations each. Generation 1 is an original pair from Experiment 1 (who did not observe anyone else's drawings before playing), Generation 1's last drawings are observed by Generation 2, Generation 2's last drawings are observed by Generation 3, and Generation 3's last drawings are observed by Generation 4.

6.2.2 Stimuli

The drawings observed in the familiarization phase. The first generation of each of the 4 chains was chosen at random from the 12 games of Experiment 1. Each subsequent generation observes what was drawn for the last 30 trials of the previous generation's game. In practice, this amounts to observing the signs for 15 – 20 (M = 18.25) of the 26 core items from the previous generation.

The items communicated in the game. These were exactly the same as in Experiment 1.

6.2.3 Procedure

Participants first read instructions. The full instructions to the game are included as Appendix E. They are identical to those for Experiment 1, except that they include information about the familiarization phase. Each participant reads that, during the familiarization phase, he and his partner will see what another team was drawing at the end of their game. Importantly, each participant reads "Seeing what another team did may or may not help you in your game. When you're the Drawer, just do whatever will get your partner to guess correctly and quickly."

The familiarization phase lasts 15 minutes. Each subject in a pair sees screenshots of what was drawn for the last 30 trials of the previous generation's game, in the order they occurred in the original game. For each screenshot/trial, the subjects read what the item was, who drew it, and what the Guesser guessed for it. The participants study each screenshot and its associated information for 20 seconds each.

The pair of participants then plays the same communication game used in Experiment 1 (again, for two hours).

6.3 Analysis

As in Experiment 1, the Initial and Final sets of signs were coded for systematic compositionality.

In addition, the observed sets were coded for systematicity. As pairs observed the last 30 trials of another pair's game, and the items occur with varying frequency and in random order, each observed set consists of several drawings for some core items and no drawings for other core items. We reasoned that the last trial observed would be most salient to the current pair, and so coded the last drawing for each item that occurred in the observed set. There were just two amendments to the coding instructions used in Experiment 1. First, if the drawings in a category shared an element that was prohibited (e.g., the cross for medical items), those drawings were excluded from the systematicity analysis. (The number of drawings was subtracted from the total possible systematicity score.) This is because the next generation was not allowed to copy a prohibited element, and here we are interested in how well one generation's systematicity persists through the next generation. This only happened once. Second, the observation bottleneck sometimes results in there being just one drawing in a category. Of course there can be no element shared across drawings in this case, so these drawings were also excluded from the systematicity analysis. This happened just six times, in just four sets of signs.

As in Experiment 1, the sets were coded blind and in random order. In addition, the full sets were coded before any of the observed subsets. This was to prevent the coder from inadvertently looking for sign elements she marked in the subsets while coding the full sets.

As reported for Experiment 1 (in Chapter 4), we established inter-coder reliability for the coding procedure.

6.4 Results

6.4.1 Systematicity increases over generations

We asked whether the systematic compositionality that emerged in Experiment 1 could be maintained through generations. In particular, do new people, faced with the task of communicating, *choose* to use the (often arbitrary) systematic compositionality they observed?

Figure 22 shows the systematic compositionality in each pair's final set of signs, organized by chain and generation. It's clear from this that systematic compositionality is being maintained over generations within a chain. In fact, it tends to increase[21].

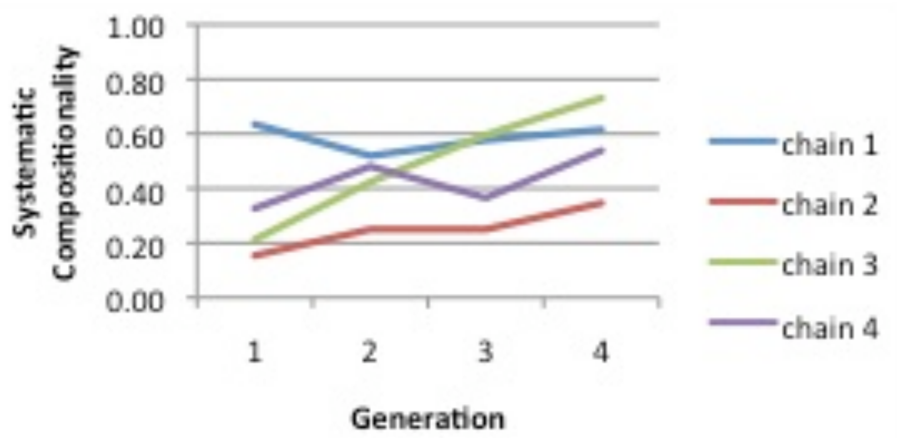


Figure 22. Systematic compositionality of final sets of signs by generation. There is a (non-significant) increase in systematic compositionality over generations.

Let's look at the chains step-by-step to determine the origin(s) of this increase in systematicity.

6.4.2 From a generation's final set of signs to the subset of it that the next generation observes

It could happen that a set of signs has less systematicity than a particular subset of it. For example, by chance, the idiosyncratic signs may not be observed. Since our systematicity measure is a proportion, this would make the systematicity of the subset greater than that of the set.

This turns out not to be the case. A Wilcoxon Signed-Rank Test found no significant difference between the systematic compositionality scores of Final sets of signs and that of their associated subsets ($p = 0.2238$).

---

[21] This increase is not significant. This may be due to the small sample sizes.

6.4.3 From the set of signs a generation observes to the first signs they produce

Instead, the increase in systematic compositionality occurs when pairs produce their Initial signs, after observing the previous generation's signs. The systematic compositionality in the Initial sets of signs pairs produced is greater than that of the set of signs they observed. ($M_{Initial} = 0.44$, $SD = 0.13$; $M_{Observed} = 0.36$, $SD = 0.2$) A Wilcoxon Signed-Rank Test confirmed this. ($p < 0.05$)

The fact that the systematic compositionality is increasing at this stage suggests that people are generalizing: they are drawing unobserved items in a way that is more systematic than the original pair drew them.

6.4.4 From a pair's initial signs to their final signs

As in Experiment 1, pairs tended to maintain the systematicity they produced in their Initial signs through to their Final signs. A Wilcoxon Signed-Rank Test found no significant difference between the amount of systematic compositionality in pairs' Initial sets of signs and that of their Final sets of signs. ($p = 0.82$, $M_{Initial} = 0.43$[22], $SD = 0.13$; $M_{Final} = 0.44$, $SD = 0.17$)

6.4.5 Summary of systematicity results

Systematicity is not only maintained over generations; it increases. The increase doesn't happen as a result of taking a subset of the signs to be observed. That is, one generation's final signs are no less systematic than the subset of them that the next generation observes. Nor does the systematic compositionality increase over the course of a generation's game - the final signs a generation produces are no more systematic than their initial signs. Rather, systematic compositionality is increased from pairs' observed subsets to their initial signs. This suggests that, with each

[22] This mean is slightly different than the mean reported in the previous paragraph because that was the mean systematic compositionality of the Initial sets of signs produced by pairs who had observed another pair's set. The current mean is the mean systematic compositionality of the Initial sets of signs produced by all pairs (including the first generation, who did not observe anyone's signs.

generation, some idiosyncratic signs (perhaps those not observed) are being replaced with systematic ones.

6.5 Transmission possibilities and issues

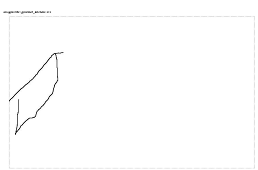
Here we offer a more qualitative analysis of when and how individual sign elements (conventions for representing semantic features) are adopted by subsequent generations.

Note that the drawings for a category of items sometimes have more than one element in common, exhibiting a family resemblance. When coding for systematicity, we focused on just the strongest shared element. Here we looked at all of the shared elements.

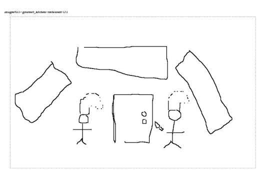
We found that subjects exhibited a wide range of behaviors.

6.5.1 Similarity of drawings to the previous generation's

Participants often draw an item the same way as they observed. Further, they are comfortably copying even pretty arbitrary drawings. (Based on their feedback, this may depend on whether they can make sense of them enough to memorize them.)

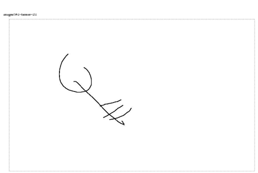


observed drawing of gourmet kitchen

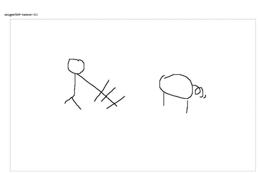


first drawing of gourmet kitchen

Sometimes this involves re-interpretation of the observed drawing. In the case below, the observed pair was drawing corn in the farmer's mouth, but the next generation understood this as a rake:

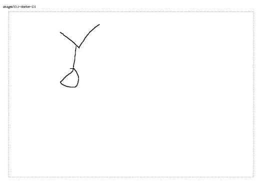


observed drawing of farmer



first drawing of farmer

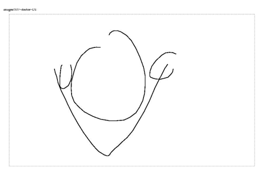
Also, one pair observed this for doctor:



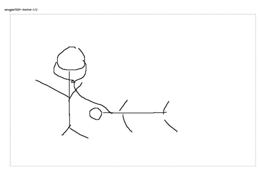
The one participant in the pair understood it as a stethoscope but the other reported after the experiment: "I just thought it was a guy upside down", that he never understood what it had to do with doctor, and but just copied from the other pair.

6.5.2 Adding detail to observed signs

Participants often add detail to the signs they observe. It could be that the Drawer remembers the sign he observed exactly but isn't sure that the Guesser will, that the Guesser doesn't guess quickly enough so the Drawer adds detail, or that what the Drawer "remembers" is not the exact drawing but his interpretation of it. Here's an example:



observed drawing of doctor

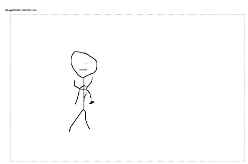


first drawing of doctor

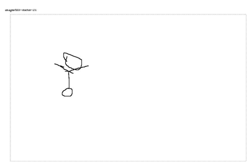
6.5.3 Cumulative simplification of signs

Cumulative simplification can be observed in one chain's initial drawings of doctor:

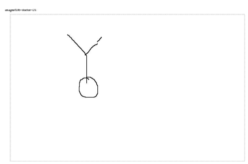
generation 1's initial drawing of doctor



generation 2's initial drawing of doctor



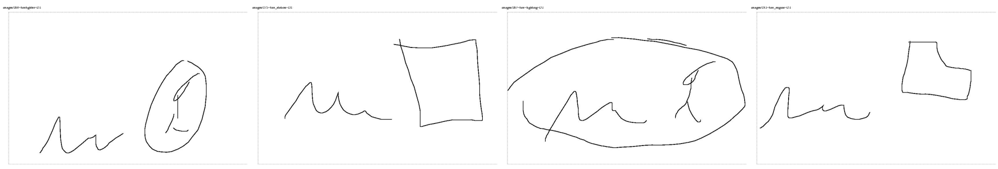
generation 3's initial drawing of doctor



generation 4's initial drawingo of doctor

6.5.4 Systematic compositionality is not always maintained.

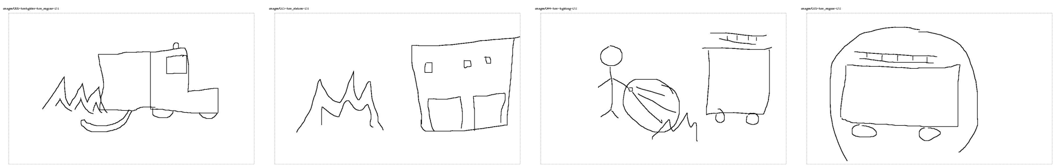
Regarding systematic compositionality, there was at least one example of a highly systematically compositional system being abandoned by the next generation. One generation's final drawings for fire items were:



firefighter fire station fire-fighting fire engine

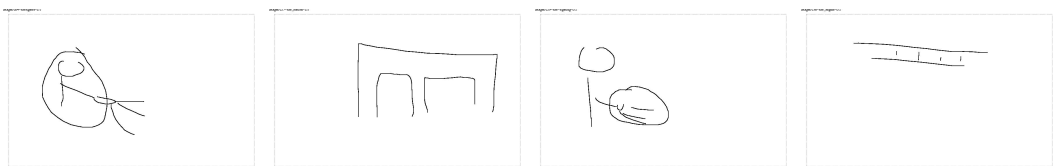
They each include the same wave-like figure on the left.

Each of these was observed by the next generation one or two times. Yet here are that generation's initial drawings for the fire items:



firefighter fire station fire-fighting fire engine

And here are their final drawings for the fire items:

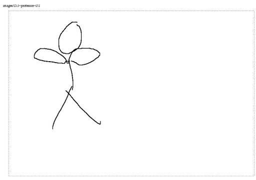


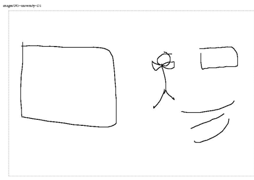
firefighter fire station fire-fighting fire engine

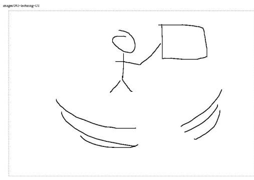
These signs are much less systematically compositional.

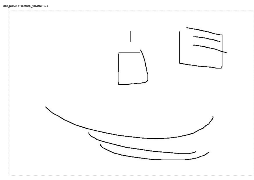
This may have been due to some combination of the participants adding detail to drawings they observed (later cutting the regular/systematic detail out, leaving just idiosyncratic elements) and them not picking up on distinctions the previous generation made.

As another example, a sign for university items emerged, was just transparent enough to be copied by the next generation, but then became too obscure to be copied by the following generation. Here are some of the drawings of university items that the third generation observed from the second:

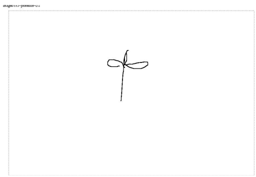

professor

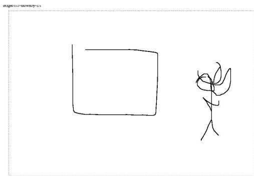

university

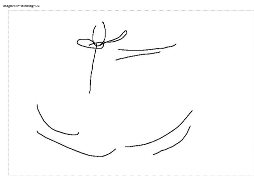

lecturing


lecture theatre

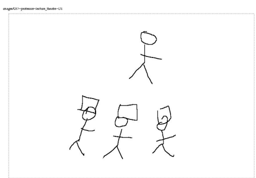
The third generation kept the bowtie. Here are some of the third generation's drawings of university items that the fourth generation observed:

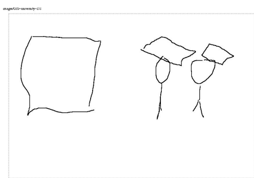

professor


university

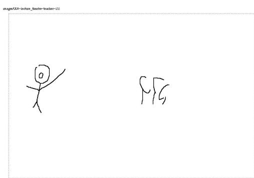

lecturing

But then the fourth generation didn't use the bowtie. After the experiment, they reported that they didn’t know what it was. Here's the fourth generation's initial drawing of each university item:

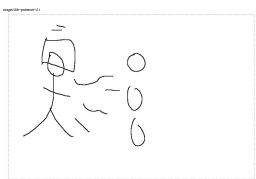

professor

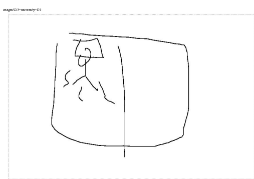

university


lecturing

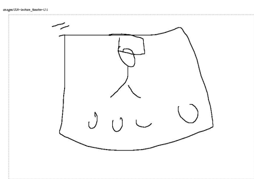

lecture theatre

And here's their final drawing of each university item:


professor


university

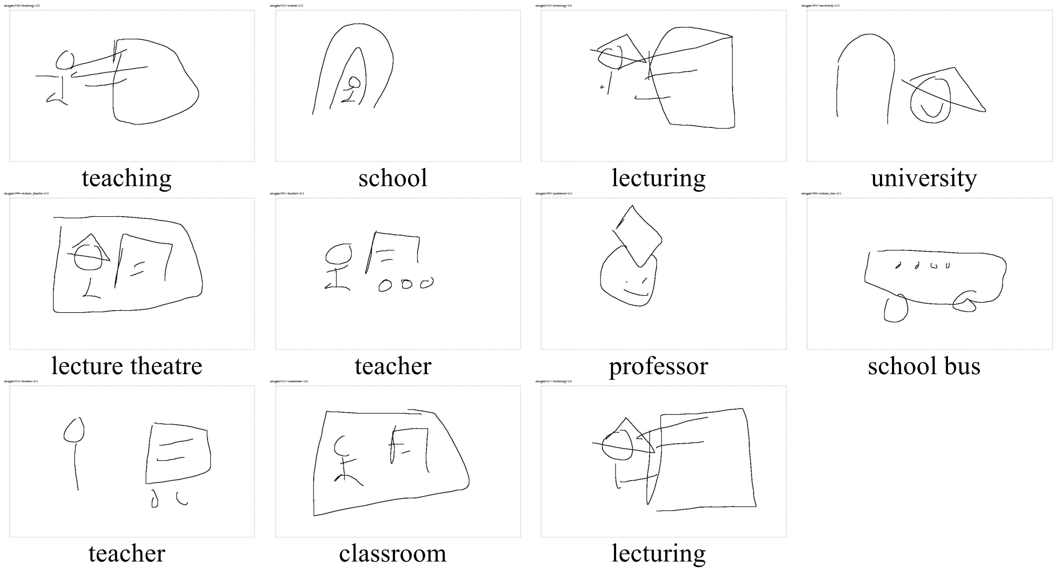

lecturing


lecture theatre

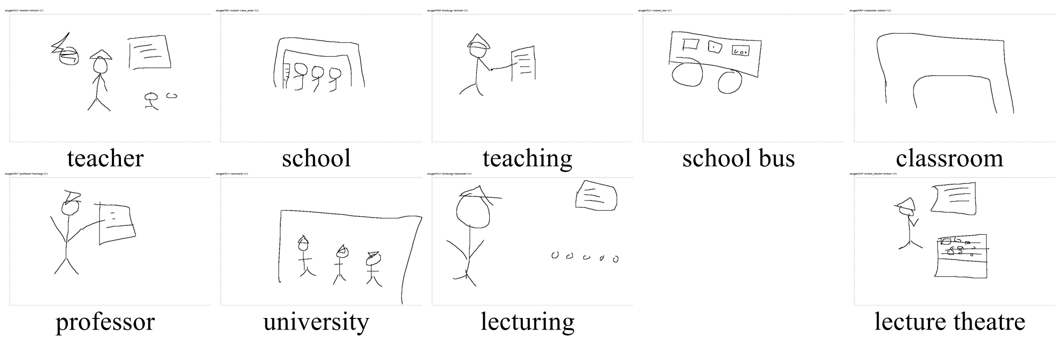
The bowtie has been completely lost.

6.5.5 Collapsing distinctions observed

Participants often don't pick up on distinctions they observed. For example, one generation was distinguishing the primary education items from the university items by drawing the professor with a mortarboard and the teacher with a chalkboard and children. Here are the examples the next generation observed:

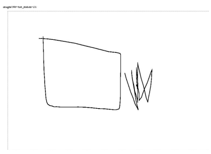


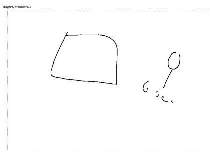
Below are the first drawings for each of the primary education and university items from the next generation. Drawings for *both* primary education and university items include the mortarboard, chalkboard, and children – the distinction between university and primary education items has been lost.

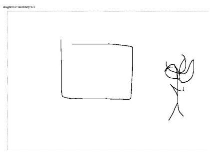


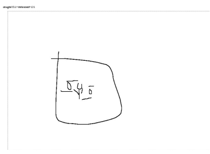
6.5.6 Increasing systematicity in unobserved items

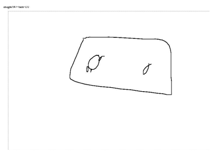
Sometimes drawings for unobserved items are fairly systematic. One pair's initial drawing of hospital is a good example. Here some of the buildings (first row) and medical items (second row) that they observed:

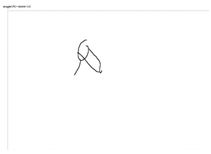

fire station

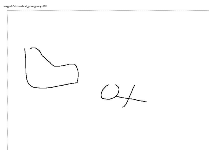

school

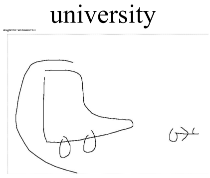

university


restaurant

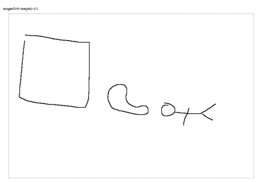

barn


doctor


medical emergency


ambulance

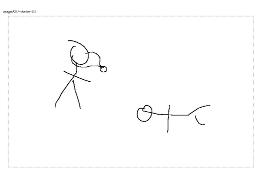
Strikingly, their initial drawing of hospital (which they had not observed) has the building component and the medical component:



It's perfectly systematically compositional.

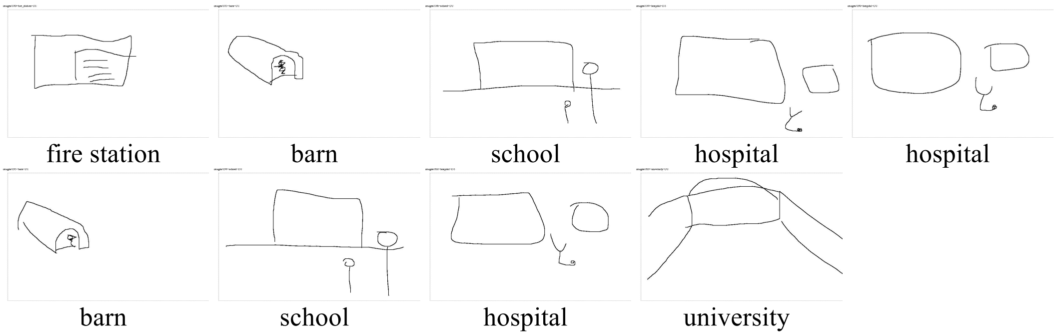
6.5.7 Increasing systematicity in observed items

Participants have changed even *observed* signs to be more regular. The pair above added the medical component to their initial drawing of doctor (which they had observed as just a person with a stethoscope):

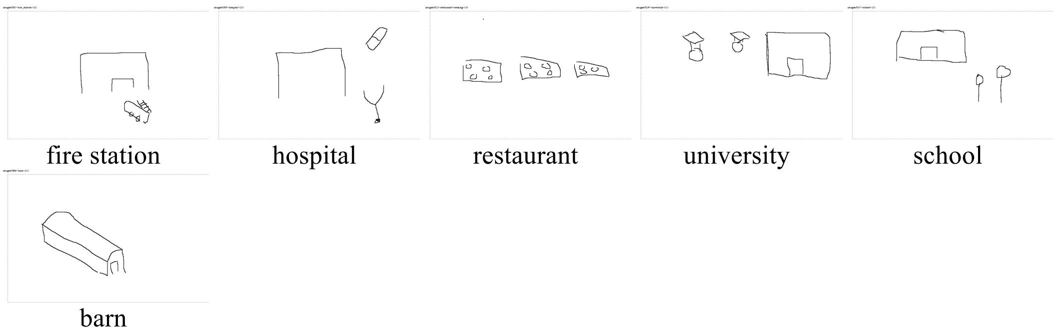


Here's another example of participants regularizing a drawing they had observed:

all drawings of buildings which one pair observed from another



that pair's initial drawings of buildings



Notice how university and fire station have changed. Even school has lost the horizon line. (Note that, as here, restaurant is often an exception – participants reported not thinking of it as a building.)

6.5.8 The role of arbitrariness

Subjects provided informative feedback about what they copied and why. Strikingly, the degree of (perceived) motivatedness of the whole signs or sign elements was often reported as a major factor.

Whether an element gets adopted or not seems to interact with arbitrariness. This is an issue that previous work on the evolution of systematicity, which started from a set of arbitrary signs, was unable to address. Under the current view, in which signs do not necessarily become totally arbitrary before systematicity evolves, we see that people have an alternative to adopting any arbitrary systematicity they observe: they may decide that a novel, motivated sign is more likely to be successful.

6.6 Summary

Thus, conventions developed in a dyad which make sets of signs systematically reuse arbitrary elements aren't lost to subsequent generations, but rather can be seen as the start to real languages which are used in communities and generations of people. When faced with the task of communicating quickly and accurately, subjects will sometimes adopt other's conventions for representing semantic features.

In addition, qualitative analysis of some of the signs produced in this experiment give us a sense of the issues that might be addressed in future work. Exploring the evolution of systematicity starting from a set of arbitrary signs has allowed us to overlook many of these issues.

Chapter 7

Conclusions

7.1 Our contribution

The aim of this thesis is to explore how the systematic re-use of arbitrary elements arises in a language.

In Chapter 1, we argued that two common ways of characterizing signs – by the degree of motivatedness they exhibit and by the degree of systematic compositionality they exhibit – are not necessarily orthogonal. Thus, the emergence of arbitrariness and of systematic compositionality in language should be studied together. We focused on a particular interface of the two properties: the systematic re-use of arbitrary elements.

In Chapter 2, we noted that previous work that demonstrates how novel signs can emerge and then evolve to become arbitrary does not measure systematic compositionality. On the other hand, previous work on systematic compositionality proposes a mechanism for the evolution of systematic compositionality and a measure of the property, but does not address the evolution of arbitrariness.

We introduced a 'parallel theory' of the emergence of the systematic re-use of arbitrary elements in Chapter 3, in which arbitrariness and systematic compositionality arise in a parallel, rather than serial, manner. We argued specifically that systematic compositionality can emerge before signs have become completely arbitrary and that this systematic compositionality can be maintained as signs become arbitrary.

In Chapter 4, we reported the first of a series of experiments that probe how the systematic re-use of arbitrary elements arises in novel communication systems. In Experiment 1, partners communicated items to each other by drawing on a whiteboard. They were prohibited from using conventional language or any pre-existing symbols, so they had to create their signs from scratch. They communicated about the same limited set of items over a long period of time. In this way, we were able to examine how their drawings for the items changed over the course of interaction. The items they communicated about were structured, i.e. they shared semantic features with each other. In this way, we could examine the degree of systematic compositionality in the sets of signs. We found that the systematic re-use of arbitrary elements emerged. Further, the evolution of arbitrariness and systematic compositionality were parallel: even participants' first drawings of items were systematically compositional, and this systematic compositionality was maintained as the signs became more arbitrary.

Chapter 5 presented Experiment 2, in which we began to explore how the systematic re-use of arbitrary elements that arises could be extended beyond pairs of closely interacting communication partners. Naïve participants were trained on a subset of a very systematically compositional set of signs and asked to guess what the original participants drew for the missing items. Despite the fact that the naïve participants played no role in – indeed, did not even observe – the creation of the sign systems, they could detect the systematic compositionality in them and generalize from it.

Chapter 6, the previous chapter, presented Experiment 3, which tested whether participants actually *do* make use of the systematic re-use of arbitrary elements that they observe in others' sign systems, when faced with the task of communicating (rather than learning and reproducing signs). Participants played the same game as in Experiment 1 but observed the end of another pair's game for 15 minutes before starting their own. Pairs of participants were organized into generations, so that a pair observed signs produced by a pair who played before them, who in turn observed signs from a pair who played before them, and so on. The systematic compositionality was not only maintained, but appears to have increased, over generations. The increase in systematic compositionality occurred when pairs created signs for items they did not observe – presumably, as they generalized using the systematic compositionality they observed.

In sum, the main contribution of this thesis has been to explore an alternative mechanism for the emergence of the systematic re-use of arbitrary elements: arbitrariness and systematic compositionality emerge in a parallel fashion within the dyad, and subsequent communicators maintain – or even increase – the structure they have observed.

7.2 Broader implications

Our results carry some broader implications. We'll examine the relevance of our work beyond the graphical modality and how our work might contribute to a more general theory of the evolution of signs.

7.2.1 Beyond the graphical modality

Our experiments were designed to demonstrate how the systematic re-use of arbitrary elements can arise in sets of signs. While our experiments used a novel communication medium (graphics), our demonstration applies to communication in general, and spoken language in particular. We hypothesized (in Chapter 3) and then demonstrated (in Chapter 4) how systematic compositionality could emerge in novel signs. Motivatedness and interaction history were identified as key factors. There is no reason to believe that either of these factors is specific to the graphical modality. Consider interaction history first, and recall the pair who drew "school bus" as a school bus next to a chalkboard because they had earlier drawn "teacher" as a stick figure next to a chalkboard (resulting in the signals for similar meanings sharing an element). We could see the same effect in the spoken modality: if a pair's vocalization for a meaning included a certain sound, one person in the pair could very well produce that same sound in the vocalization for a related concept later, as a way of referencing the first. As for the second factor, motivatedness, recall our various demonstrations of how the beginnings of spoken language could have been motivated (in Chapter 2). To name just a few, some sounds carry meaning cross-linguistically, some sound-meaning correspondences that appear in one language are actually recognized universally, and experiments show that people associate many different aspects of sound with many different aspects of meaning.

Interestingly, along the way we have also discovered some clues about how spoken language might have gotten off the ground. First, we have shown that there are many motivated ways to refer to a given item. We discussed this while reviewing previous work on how the beginnings of spoken language could be motivated in Chapter 2. We then found in our experiments that, for a given meaning, there are many signals that have some kind of inherent connection to it. Recall, for example, the great many associations that participants made to the item "farmer", which we illustrated in our discussion of Experiment 1 in Chapter 4. This lends additional support to the idea that a speech signal need not sound like the actual meaning, but rather could sound like some association to the meaning.

Second, it seems that motivatedness is crucial only for the very first signs. Once several signs have been established, new signs can be built off of those. While presenting our hypothesis in Chapter 3, we proposed that initial systematic compositionality in signs could be due to the interaction history shared by partners. Experiment 1 confirmed that interaction history was indeed a very important factor: the systematic compositionality we see in initial sets of signs is specific to the pairs. There seems to be a tendency of communication partners to re-use elements from signs they already have when creating new signs. Thus, people creating a communication system don't need to find a motivated way of expressing every meaning.

Considering all the ways sound can map to meaning and how small a role motivatedness need play in the creation of signs, we can see how spoken language might have evolved. Thus, although we have demonstrated the emergence of the systematic re-use of arbitrary elements in the graphical modality, our results apply to communication in general, including spoken language.

7.2.2 A more general theory of the evolution of signs

We designed an experimental paradigm that allowed us to explore the emergence of arbitrariness and the emergence of systematic compositionality together, and found that the systematic re-use of arbitrary elements can emerge in a way not previously suggested. Looking at the emergence of the two properties together was informative because, based on previous work, one might have assumed that novel signs are not systematically compositional or that, as signs become arbitrary, they would lose any systematic compositionality they had.

The broader implication of this finding is that there may be much to be gained from examining the evolution of linguistic properties *together*. Now that significant work has been done to establish how linguistic properties evolve independently, the field of evolutionary linguistics may be ready to explore interactions between them. As just one example, we might also explore how the emergence of combinatoriality interacts with the emergence of arbitrariness.

Experiments like those could contribute to a more general theory of the evolution of signs. It might predict what constrains the signal that one produces for a given meaning (in a particular context). For example, one might aim to produce a signal that resembles the meaning (motivatedness), that resembles the signal previously produced by one's communication partner for that meaning (interaction history), and/or that fits in with other signs one uses (systematic compositionality). In our experiments, we have witnessed these three particular constraints at work, but there are surely others. The relative weighting of these constraints may depend on what's in the common ground between the people communicating. For example, perhaps the only things someone can expect to share with a stranger is universal information. If so, motivatedness may be the strongest constraint. Between members of the same speech community, systematic compositionality may carry more weight.

7.3 Future Work

Our results have raised some important issues, which the experimental paradigm we presented in this thesis could be straightforwardly extended to address. The issues cluster around the emergence of the systematic re-use of arbitrary elements and the maintenance of systematic compositionality as a set of signs developed by some people is transmitted to others.

7.3.1 Emergence of the systematic re-use of arbitrary elements

Now that we have demonstrated an additional route to the systematic re-use of arbitrary elements, the obvious question is: are there others? That is, how else might the systematic re-use of arbitrary elements emerge in a language? Let's compare the most influential human experiment on the evolution of the systematic re-use of arbitrary elements, Kirby, et al. (2008)'s, with ours. We'll see they differ in a number of ways: the role of a bottleneck in the transmission of signs, how compelled the participants feel to copy others' signs (as opposed to creating their own), and the degree of feedback the participants receive about whether their signs were understood. We'll then suggest experiments that could bridge the gap.

What role need a bottleneck play in the emergence of systematic compositionality? In Kirby, et al. (2008), the participant representing the first generation learned some of the signs of an alien language and then was asked to recall all of them, the second generation participant learned some of the signs the first participant produced and then was asked to recall all of them, and so on. Thus there was a transmission bottleneck – not all of the signs were transmitted from one generation to the next[23]. This bottleneck created systematic compositionality – the initial alien languages consisted of arbitrary signs (which exhibited only chance recurrences between signal elements and meaning elements), while the final generations' languages were highly systematically compositional. This happened because each generation replaced some of the previous generation's arbitrary signs –

[23] There can also be a memory bottleneck, in which each participant is exposed to all of another participant's signs but cannot retain all of them in his memory.

presumably, those they did not observe – with signs that followed rules they noticed in the signs they *did* learn (viz. that some signals for similar meanings shared elements).

In contrast, there was no such bottleneck in Experiment 1. There was no language that the participants were supposed to learn but were only partially exposed to – our participants weren't exposed to any language; they knew they had to *create* the signs they would use. Despite the lack of a bottleneck in Experiment 1[24], systematic compositionality emerged.

Our Experiment 3 began to bridge the gap between Experiment 1 and Kirby, et al. (2008), showing how the mechanisms for the creation of systematic compositionality in each can work in tandem. Pairs of participants in Experiment 3 were tasked with communicating with each other (as in Experiment 1) but both observed another pair's signs for some of the items they would communicate. Over generations of this, the systematic compositionality in novel signs was maintained or increased.

More work is needed to determine exactly what role a bottleneck need play in the creation of systematic compositionality.

When do people communicate by copying others' signs, as opposed to creating their own?

In our Experiments 1, the task of the participants is to communicate items to their partner. They are prohibited from using any pre-existing symbols, and so they create motivated signs. The motivatedness appeared to aid the emergence of systematic compositionality.

Kirby et al. (2008)'s experiment stands in contrast. The task of their participants was to learn and then recall someone else's signs for items. Systematic compositionality

---

[24] There appeared to be no memory bottleneck in the game either. Pairs did not appear to forget how they draw any of their items. A programming error resulted in one of the filler items appearing twice in one pair's game, the second occurrence 44 trials after the first. The second drawing of that item was strikingly similar to the first, suggesting that the Drawer remembered how they had drawn the item the first time.

emerged as the participants based their guesses about signs they didn't learn on signs they did learn.

These two tasks – communicating with someone vs. reproducing someone's signs – are actually quite different. Today, with conventional language, the best way to communicate with someone is very often to use the language one copied from others. But before a community shares linguistic conventions, the pressure to communicate may often have been in conflict with the pressure to reproduce others' signs. Imagine a language is just getting off the ground, and Adam has learned signs from (or co-created signs with) Bob. Adam cannot be sure that Chris has also learned these signs. Assuming the signs that Adam learned have lost some transparency (due to simplification), the signs best for communication with Chris are not necessarily the signs he can reproduce from Bob.

So we ask: what if the task of Kirby et al. (2008)'s participants had been to communicate any way they liked? Experiment 3 moved us in the direction of addressing this, as participants in that observed others' signs (although they were not instructed to copy them). Those participants, given the task of communication, sometimes copied signs, but sometimes either created novel motivated signs or made the signs they copied more motivated. Even in Experiment 2, when the participants were tasked with guessing what a pair drew for some items (i.e. not with communicating), motivatedness played a role in their guesses.

While there's surely a strong pressure to copy signs from others once a language is established, that pressure may not be as strong from the very first signs. When people are able (i.e. allowed) to create novel, motivated signs, they may not copy others' arbitrary signs. Thus, it's an open question when the best way to communicate is by copying others' signs, as opposed to creating your own signs.

What role does the level of feedback from one's communication partner play?
In Experiment 1, partners communicated the same items to each other over a long period of time. When a participant was the Drawer, he saw what his partner guessed. When he was the Guesser, he learned what the actual item was that trial.

Kirby, et al. (2008)'s participants interacted much less directly. Each participant learned the signs produced by the previous participant, was instructed to reproduce the previous participant's signs, but did not learn whether or not he was successful. We know that systematic compositionality emerges under these circumstances, but what about arbitrariness? The initial languages in these experiments are arbitrary, so the experiments don't address the emergence of arbitrariness. In (Garrod, et al., 2007)'s experiments, signs became arbitrary only in the conditions in which participants received feedback about the success of their signs. However, it's possible that there are other pressures towards arbitrariness. Future work could probe what effect these different levels of feedback have on the emergence of the systematic re-use of arbitrary elements.

Extensions to our experimental paradigm could answer the questions outlined above. With Experiment 1, we offer a novel experimental paradigm to investigate how and when systematic compositionality and arbitrariness emerge in novel communication systems. This could be extended to address the above issues. Here we present three possible experiments. We name these the "transmission chain", "replacement", and "closed-group" experiments, borrowing Mesoudi & Whiten (2008)'s classification of methods of cultural transmission.

The transmission chain experiment. Essentially, this is a graphical communication version of Kirby, et al. (2008)'s experiment. The first participant draws the (26 core) items used in our experiments on a whiteboard. To induce the move towards arbitrariness (that signs make over a longer period of time than an experiment can capture), he is instructed to draw as quickly as he can. He is told to draw the items for someone to later identify. The next participant learns *some* of the first participant's signs, e.g., the drawings for 13 of the 26 items, selected at random. This second participant then draws all of the items for the first participant to identify, as quickly as she can. The third participant learns some of the second participant's signs (a different, randomly-chosen subset) and then draws all of them, for the second participant to identify, as quickly as he can, and so on.

This experiment is similar to Kirby, et al. (2008)'s in that there is a transmission bottleneck – a participant doesn't learn a drawing for every item he will later have to express – and in that the participant producing the signals does not receive feedback on whether he successfully communicated. However, in terms of whether one communicates by copying others' signs, this experiment is somewhere in between Kirby, et al. (2008)'s and ours. The task of the participant is to communicate to the previous participant, not to reproduce the previous participant's signs. However, drawing items the way the previous participant did may be precisely the best strategy for communicating to that participant. Thus, the main motivation for this experiment would be to test the role of feedback in the emergence of the systematic re-use of *arbitrary* elements.

The replacement experiment. The replacement experiment begins with two people (A and B) playing the graphical communication game introduced in our Experiment 1 (communicating the same set of items to each other using the whiteboard, switching roles often, receiving immediate feedback about their communicative success, etc.). However, just before the end of A and B's game, a new player (C) begins to observe them play[25]. After a while, C replaces A, and B and C continue the game. A new player (D) eventually begins to observe them and then later replaces B so that C and D are playing, and so on.

How could this experiment address our open questions? The design is very similar to our Experiment 3. In particular, there is a bottleneck on transmission of a set of signs from one pair to the next, as the observer will not observe and remember all of the pair's signs. Further, participants receive immediate feedback about their communicative success. This should allow sign elements to become more arbitrary.

The point of this experiment, thus, would be to address our question about when people decide to communicate by copying others' signs as opposed to creating novel,

[25] Participants should be told that they may be observed and may or may not swap partners during the game. In this way, a participant won't be confused when his partner is swapped, but participants also won't try to design a set of signs that would be useful with strangers.

motivated signs. The task is to communicate, not explicitly to copy someone else's signs. The pressure to copy someone else's signs is greater than in our Experiment 3, as all but the first participants will begin the game playing with someone who already has signs, and so it would be very useful to simply copy his signs. The pressure to copy someone else's signs is less than in Kirby, et al. (2008), though. First, the participant has a greater sense of co-creating the set of signs.[26] Second, the participant is generally freer to draw what he wants and simply add information if his partner does not guess from that.

The closed-group experiment. In the closed-group experiment, participants always play in pairs, and their task is exactly as in Experiment 1. The difference is that participants switch partners each round (say, six trials). There could be, for example, a community of four participants in each run of this experiment.

This experiment would explore the role of the bottleneck in the emergence of the systematic re-use of arbitrary elements. The bottleneck introduced would be a memory bottleneck – a participant may forget how to draw a given item with a given partner (viz. how the two participants have drawn that item with each other in the past). This may result in increasing systematic compositionality in the set of signs because idiosyncratic signs are less likely to be remembered than elements used in multiple signs. On the other hand, the participants may simply assume that, since they are all playing with each other, they are all using the same signs. In that case, there'd be no more of a bottleneck than in our original game.

This closed-group experiment would additionally explore the pressure to copy others' signs in an interesting way. Since players would switch partners often, there may be a pressure on the signs to retain a degree of motivatedness (Fay, Garrod, MacLeod, Lee, & Oberlander, 2004). This is, again, because one cannot be sure that

---

[26] The experimenter could further manipulate the pressure to copy vs. create novel, motivated signs by changing which of the players draws each item first. If the retained player draws a given item first, he may simply draw what he had been drawing with his previous partner. However, if the new player draws a given item first, and he has not observed or does not remember how that item had been drawn, he is likely to re-invent the sign for that item.

his current partner will understand a certain sign, because he cannot remember if they used it together, so he will err on the side of transparency. On the other hand, being part of a community may provide a pressure to copy the signs everyone else is using, whether or not they are transparent.

In fact, we have piloted the three possible future experiments just outlined. The basic designs are sound, but there are logistical issues to resolve. For example, because participants interact with many new people in these, the signs don't appear to simplify as quickly. Thus, the game time should be extended.

With these three experiments we can begin to explore the factors governing the emergence of the systematic re-use of arbitrary elements. One can imagine many other straightforward – yet highly informative – extensions to the experimental paradigm we have presented in this thesis.

7.3.2 Transmission of systematic compositionality

In addition to exploring other routes to the emergence of the systematic re-use of arbitrary elements, future work should probe deeper into the transmission of sets of signs that exhibit this property.

We began this thesis by arguing that there is good reason to explore the evolution of arbitrariness and the evolution of systematic compositionality together. Indeed, in our experiments, we consistently found that the degree of residual motivatedness of signs affected the transmission and maintenance of systematic compositionality in them. For example, in Experiment 2, participants sometimes reported using their own understanding of the drawings they learned, and of how they would draw the items themselves, in making guesses about how another pair drew the items. Thus, our experiments have demonstrated that the relationship between motivatedness and systematic compositionality is far from straightforward.

Future work could pit motivatedness against systematic compositionality in order to gain a better understanding of how they interact. We suggest an experiment that we name the "anti-motivated but systematically compositional mapping" experiment.

The anti-motivated but systematically compositional mapping experiment.
In this experiment, a participant learns a highly systematically compositional set of signs, taken from the end of a pair's game. The catch is that drawings are mapped to the items in the wrong way, e.g., the original pair's drawing for "lecture theatre" is said to be what they drew for "fire-fighting". This is much like the Random set used in Experiment 2, except that the drawings are mapped to the items in a way that preserves systematic compositionality. For example, all of the original pair's drawings for rooms are said to be their drawings for activities, and all of the original pair's drawings for university items are said to be for items relating to fire-fighting. Figure 23 shows how an anti-motivated but systematically compositional set of signs might look.

As in Experiment 2, the participants' task is to learn most of the signs in the set and then to guess what the original players drew for the missing items. To the extent that residual motivatedness plays a role in the transmission and maintenance of systematic compositionality, the participants will make the wrong generalizations.

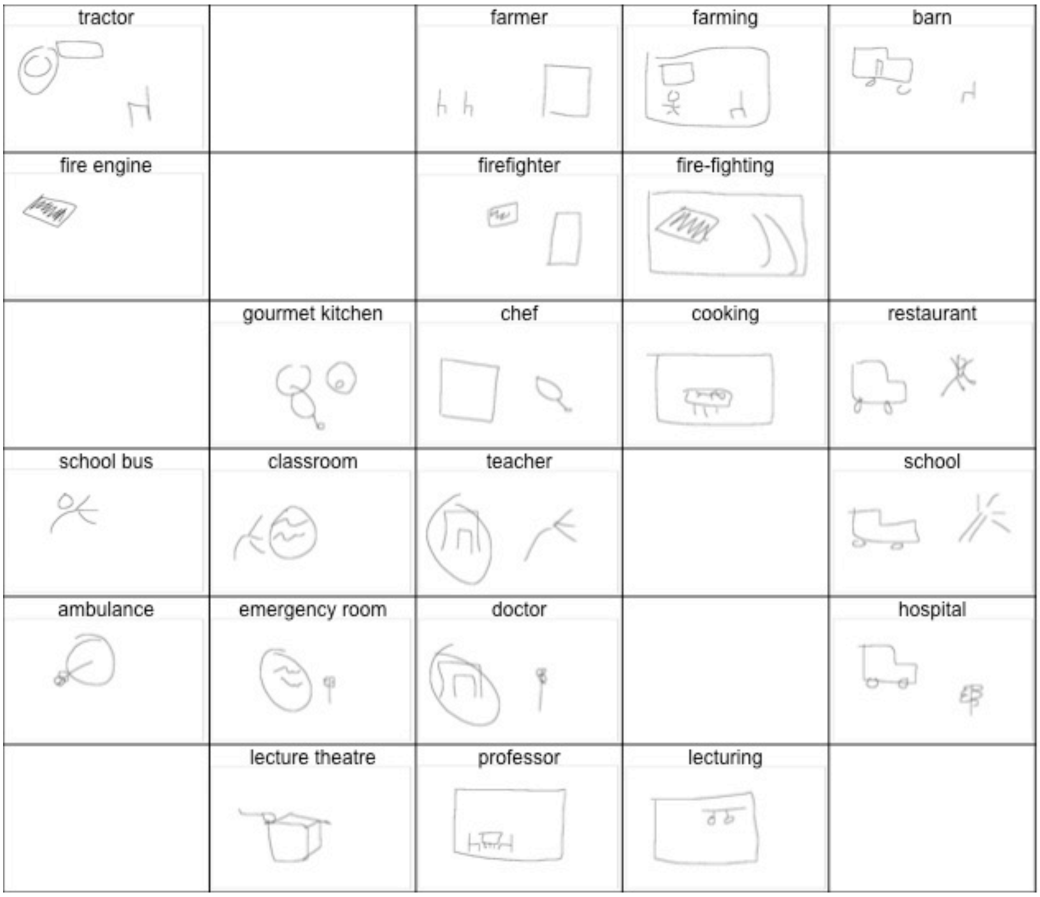


Figure 23. An example of an anti-motivated yet systematically compositional set of signs. The drawings are one pair's final signs for the items, but mapped to the wrong items. This anti-motivated mapping preserves systematic compositionality, though, in that the drawings for one semantic feature are said to be for another.

A pilot of this experiment produced interesting results. The participant was able to correctly generalize in some cases, but not all. In this, and similar pilots, participants reported seeing the iconicity in signals mapped incorrectly to their meanings. However, some incorrect mappings were rejected. For example, the pilot participant could not accept that a drawing of what was obviously a vehicle could be someone's sign for, say, building, even though she could see that it always occurred in signs for buildings and never occurred in signs for vehicles. Thus, while a signal element can be arbitrarily mapped to a semantic feature (as they are in morphology), there appear to be constraints on the form of the signal element, viz. it cannot have a strong natural connection to a different meaning. This is just one example of ways we might see motivatedness and systematic compositionality interact.

7.4 Concluding remarks

In sum, we hope that this thesis is just the beginning of an interesting and important line of research.

Bibliography

Ben-Artzi, E., & Marks, L. (1999). Processing linguistic and perceptual dimensions of speech: Interactions in speeded classification. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 579-595.

Bickerton, D. (2003). Symbol and structure: a comprehensive framework for language evolution. *Language evolution*, 77-93.

Bloomfield, L. (1933). *Language*. New York: Allen & Unwin.

Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(6), 1482-1493.

Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews, 2*(3), 177-226.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*(4), 711.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics, 37*(4), 575-596.

Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1-39.

Croft, W. (2005). Logical and typological arguments for Radical Construction Grammar. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive grounding and theoretical extensions* (pp. 273-314): John Benjamins Publishing Company.

De Ruiter, J., Noordzij, M., Newman-Norlund, S., Hagoort, P., & Toni, I. (2007). On the origin of intentions. In P. Haggard, Y. Rossetti & M. Kawato (Eds.), *Attention & Performance XXII: Sensorimotor foundation of higher cognition* (pp. 593-610). Oxford: Oxford University Press.

Fay, N., Garrod, S., MacLeod, T., Lee, J., & Oberlander, J. (2004). *Design, adaptation and convention: The emergence of higher order graphical representations*.

Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1509), 3553-3561.

Fisher, C., Gleitman, H., & Gleitman, L. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology, 23*(3), 331-392.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Connections and symbols*, 3-71.

Frishberg, N. (1975). Arbitrariness and iconicity: historical change in American Sign Language. *Language*, 696-719.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science: A Multidisciplinary Journal, 29*(5), 737-767.

Garcia-Alvarez, I. (2005). Gottlob Frege. In S. Chapman & C. Routledge (Eds.), *Key thinkers in linguistics and the philosophy of language*. New York, New York: Oxford University Press.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*(2), 181-218.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition, 53*(3), 181.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science: A Multidisciplinary Journal, 31*(6), 961-987.

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*: University of Chicago Press.

Goldberg, A. (2003). Constructions: a new theoretical approach to language. *Trends in cognitive sciences, 7*(5), 219-224.

Goldin-Meadow, S., Mylander, C., & Butcher, C. (1995). The resilience of combinatorial structure at the word level: Morphology in self-styled gesture systems. *Cognition, 56*(3), 195-262.

Grice, H. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts* (pp. 41-58). New York: Academic Press.

Grzega, J., & Schoener, M. (2007). *English and General Historical Lexicology*. Ingolstadt: Katholische Universitaet Eichstaett.

Haiman, J. (1985). *Natural Syntax*: Cambridge: Cambridge University Press.

Healey, P., Garrod, S., Fay, N., Lee, J., & Oberlander, J. (2002). *Interactional context in graphical communication.* Paper presented at the Proceedings of the 24th Annual Conference of the Cognitive Science Society.

Healey, P., Swoboda, N., & King, J. (2002). A tool for performing and analysing experiments on graphical communication. *People and Computers*, 55-68.

Healey, P., Swoboda, N., Umata, I., & King, J. (2007). Graphical Language Games: Interactional constraints on representational form. *Cognitive Science: A Multidisciplinary Journal, 31*(2), 285-309.

Hockett, C. (1960). The origin of speech. *Scientific American, 203*(3), 88-96.

Hurford, J. (2000). The emergence of syntax. *The evolutionary emergence of language: Social function and the origins of linguistic form*, 219-230.

Hurford, J. (2004). Human uniqueness, learned symbols and recursive thought. *European Review, 12*(04), 551-565.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54-65.

Iwasaki, N., Vinson, D., & Vigliocco, G. (2007). What do English Speakers Know about gera-gera and yota-yota?: A Cross-linguistic Investigation of Mimetic Words for Laughing and Walking. *Japanese Education in the World, 17*, 53-78.

Kalish, M., Griffiths, T., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review, 14*(2), 288.

Keller, R. (1998). *A theory of linguistic signs* (K. Duenwald, Trans.): Oxford University Press New York.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. *The evolutionary emergence of language: Social function and the origins of linguistic form*, 303-323.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation, 5*(2), 102-110.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681.

Klima, E., & Bellugi, U. (1979). *The signs of language*: Harvard Univ Pr.

Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: verb learning and the footprint of universal grammar. *Cognition, 87*(3), 151-178.

Marks, L. (1987). On Cross-Modal Similarity: Auditory-Visual Interactions in Speeded Discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 384-394.

Marks, L. E. (1974). On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness. *The American Journal of Psychology, 87*(1/2), 173-188.

Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1509), 3489.

Monaghan, P., & Christiansen, M. (2006a). Iconic versus arbitrary mappings and the cultural transmission of language. In A. Cangelosi, A. Smith & K. Smith (Eds.), *The evolution of language* (pp. 430-432). New Jersey: World Scientific Publishing Company.

Monaghan, P., & Christiansen, M. (2006b). *Why Form-Meaning Mappings are not Entirely Arbitrary in Language.* Paper presented at the CogSci.

Nuckolls, J. B. (1999). The Case for Sound Symbolism. *Annual Review of Anthropology, 28*, 225-252.

Parault, S., & Schwanenflugel, P. (2006). Sound-symbolism: A Piece in the Puzzle of Word Learning. *Journal of psycholinguistic research, 35*(4), 329-351.

Peirce, C. (1955). *Philosophical writings of Peirce*: Courier Dover Publications.

Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(02), 169-190.

Ramachandran, V., & Hubbard, E. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3-34.

Reilly, J., Biun, D., Cowles, H., & Peelle, J. (2008). Where did Words Come from? A Linking Theory of Sound Symbolism and Natural Language Evolution.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences, 102*(7), 2661-2665.

Saussure, F. (1916). *Course in general linguistics*. New York: McGraw-Hill.

Schober, M., & Clark, H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*(2), 211-232.

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. S. (2009). Signalling signalhood and the emergence of communication. *Cognition, 113*(2), 226-233.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences, 104*(18), 7361.

Simner, J., Cuskley, C., & Kirby, S. (under review). What Sound does that Taste? Cross-modal Mappings Across Gustation and Audition

Smith, A. (2008). Protolanguage reconstructed. *Interaction Studies, 9*(1), 100-116.

Smith, K. (2006). The Protolanguage Debate: Bridging the Gap? In A. Cangelosi, A. Smith & K. Smith (Eds.), *The Evolution of Language: Proceedings of the 6th International Conference* (pp. 315): World Scientific Publishing Company.

Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua, 117*(3), 579-604.

Tamariz, M., & Smith, A. (2008). *Regularity in Mappings Between Signals and Meanings*.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication, 18*(1), 47-67.

www.writtensound.com. Retrieved June 7, 2009, from http://www.writtensound.com

Appendices

Appendix A: Instructions to participants in Experiment 1.

You two are going to communicate items to each other by drawing. You'll take turns being the Drawer and the Guesser.

When you're the Drawer, you'll see an item (such as Arnold Schwarzenegger) at the top of your screen. You want your partner to guess exactly that. You may NOT use any symbols or drawing conventions, such as:

- letters
- numbers
- punctuation (?, !, etc.)
- mathematical signs (+, =, <, etc.)
- pointing with arrows
- universal signs (the cross for first aid, a skull and cross-bones for danger, etc.)
- drawing conventions (a heart to represent love, + over the eyes to show that someone is dead, etc.)

Imagine your partner is an alien who has spent time on Earth but has never drawn, written, or read anything before. If the alien wouldn't understand it, it's not allowed!

When you're the Guesser, as soon as you can, guess by typing into the chat window. You may not type anything other than your actual guess. (For example, you may NOT type *I think it's Arnold Schwarzenegger* or *Arnold Schwarzenegger?*) You may not scroll up in your chat window to remind yourself of previous guesses. Your team gets just one guess per item.

As the Drawer, you must stop drawing immediately. Type exactly what your item was, and nothing else. (For example, you may NOT type *Right!* or *You were close - it was Arnold Schwarzenegger.*)

Your team gets 1 point for every correct guess and loses 1 point for any incorrect guess or drawing that included a symbol or convention. Your team's goal is to get as many points as possible in the allotted time. Subjects from the top three teams will be entered into a prize draw for an additional 20 GBP!

The items are generated randomly. You may get the same item twice, even in a row.

Appendix B: Instructions for coding systematic compositionality.

You're going to see sets of drawings from a communication game. In it, pairs of people communicated items (such as "teacher" or "medical emergency") to each other by drawing on a whiteboard. Partners played with each other for a very long time, drawing the same 26 items for each other over and over again.

You're going to see sets of drawings. Each set contains a drawing for each of the 26 items (below), and represents what players were drawing at a point in time in their game.

	people	*activites / situations*	*buildings*	*rooms*	*vehicles*
primary education	teacher	teaching	school	classroom	school bus
university education	professor	lecturing	university	lecture theatre	
medical emergencies	doctor	medical emergency	hospital	emergency room	ambulance
fire-fighting	firefighter	fire-fighting	fire station		fire engine
farming	farmer	farming	barn		tractor
gourmet cooking	chef	cooking	restaurant	gourmet kitchen	

I want to know how systematic each set of signs is, i.e. to what extent drawings for similar items have an element in common. Notice that the drawings in each row and in each column will be for similar items. For example, the drawings in the fourth row of each set will be for items that relate to fire-fighting. You're going to look at each row and each column and tell me if any of the drawings in it have an element in common. If so, briefly describe or draw the element and mark which of the drawings have it.

Only pick out an element if you're sure that there was a special understanding between the players to draw a certain thing (or things) a certain way to mean that those items relate to fire-fighting, that those items are buildings, etc.

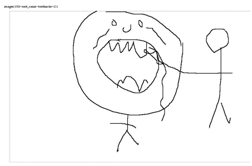
It's completely fine if the drawings in a row or a column have nothing at all in common. You may even see totally unsystematic sets of signs, in which none of the rows or columns have anything in common.

Appendix C: Instructions to participants in Experiment 2.

the signs

Two people played a game where they had to communicate items to each other by drawing on a whiteboard. It was sort of like Pictionary, except they communicated the same 40 items over and over to each other, for about 2 hours. They took turns drawing and guessing.

I'm going to show you some of the signs these people were using at the very end of their game. You'll see a screenshot of the whiteboard at the moment the guesser guessed. You'll see what the item was (e.g., “root canal”) and which player was drawing (either 1 or 2) underneath the screenshot, like this:



root_canal-D1.jpg

You'll only see signs that were guessed correctly. (The above “root canal” was not!)

your task

You have 15 minutes to learn these signs.

After your 15 minutes, you'll be tested - you'll have to draw these players' signs for some of the items. For example, I might give you “root canal” and you'd have to draw the sign above. The better you copy their signs, the better your chances of winning the 20 pound prize draw!

Look at the signs as many times, for however long, and in any order you want. You may want to copy the signs (in the Paint window), to practice drawing them.

While you're learning the signs, note:

- You're seeing only *some* of their signs - they had to communicate about more items than you see.
- During their game, some of the items came up very often and others not as

often.
- Some of the signs you'll see are from player 1 and some from player 2. You may notice slight differences in their styles. (For your test, you can draw the items either way.)
- You'll see the signs in a different order than they appeared to them in their game.
- A player stopped drawing the moment his partner guessed. Some of the signs you're seeing might be interruptions.

Ready?

When you're ready, start the timer and start looking through the signs. *After* your 15 minutes is up, scroll down for the test instructions.

[on the next page]

the test

your task

Your task now is to draw what these people drew for some items.

Recall that you saw only some of their signs - they had to communicate about more items than you saw. You're going to draw items you *didn't* see. I will compare your guess with their actual sign. The closer the match, the better your chances of winning the 20 pound prize draw!

You can look at the drawings you just studied when making your guess.

If you can, please explain your guesses to me.

Take your time!

You only have a handful of items to draw. That means you have several minutes to draw each. It's important for my research - and your chances in the prize draw - that you make your very best guess for each item.

Appendix D: Instructions to similarity judges in Experiment 2.

You have two sets of strips of drawings (B & C). Do the following for one set at a time:

1. Organize the 12 strips into 6 pairs of strips. Group strips together if they are more similar to each other than they are to any other strips.

   That is, make strip X and strip Y a pair if strip X is more similar to strip Y than it is to any other strip, and strip Y is more similar to strip X than it is to any other strip.

   It may be hard to decide. Just do your best to pair the strips according to similarity.

2. Write down your 6 pairs of strips.

3. For each of your 6 pairs of strips, rate how similar they are to each other on a scale of 1 to 10. 1 means there's no similarity at all between the drawings in the strips. 10 means the drawings in the strips are exactly the same.

Now repeat this for the other set of strips.

Appendix E: Instructions to participants in Experiment 3.

how to play the game

You two are going to communicate items to each other by drawing. You'll take turns being the Drawer and the Guesser.

When you're the Drawer, you'll see an item (such as Arnold Schwarzenegger) at the top of your screen. You want your partner to guess exactly that. You may NOT use any symbols or drawing conventions, such as:

- letters
- numbers
- punctuation (?, !, etc.)
- mathematical signs (+, =, <, etc.)
- pointing with arrows
- universal signs (the cross for first aid, a skull and cross-bones for danger, etc.)
- drawing conventions (a heart to represent love, + over the eyes to show that someone is dead, etc.)

Imagine your partner is an alien who has spent time on Earth but has never drawn, written, or read anything before. If the alien wouldn't understand it, it's not allowed!

When you're the Guesser, as soon as you can, guess by typing into the chat window. You may not type anything other than your actual guess. (For example, you may NOT type *I think it's Arnold Schwarzenegger* or *Arnold Schwarzenegger?*) You may not scroll up in your chat window to remind yourself of previous guesses. Your team gets just one guess per item.

As the Drawer, you must stop drawing immediately. Type exactly what your item was, and nothing else. (For example, you may NOT type *Right!* or *You were close - it was Arnold Schwarzenegger.*)

Your team gets 1 point for every correct guess and loses 1 point for any incorrect guess or drawing that included a symbol or convention. Your team's goal is to get as many points as possible in the allotted time. Subjects from the top three teams will be entered into a prize draw for an additional 20 GBP!

The items are generated randomly. You may get the same item twice, even in a row.

familiarization

Before you play your game, I'll show you and your partner what another team was doing at the very end of their game.

For each of the last 30 items of their game, you'll see a screenshot of the whiteboard at the moment the Guesser guessed.

Remember that the Guesser guessed as soon as possible - possibly "interrupting" the Drawer.

At the top of each screenshot, you'll read:

- who was the Drawer (Player 1 or Player 2),
- what item the Drawer was trying to communicate, and
- if the guess was incorrect, what the Guesser guessed. (If you don't read this, that means the Guesser guessed correctly.)

Seeing what another team did may or may not help you in your game. When you're the Drawer, just do whatever will get your partner to guess correctly and quickly.

Appendix F: Publications

In this section I include two publications:

Theisen, C., Oberlander, J., & Kirby, S. (2009). Systematicity and Arbitrariness in Novel Communication Systems. *Proceedings of CogSci 2009*.

Theisen, C., Oberlander, J., & Kirby, S. (in press). Systematicity and arbitrariness in novel communication systems. *Interaction Studies*.

Systematicity and Arbitrariness in Novel Communication Systems

Carrie Ann Theisen (C.A.Theisen@sms.ed.ac.uk)
School of Informatics, University of Edinburgh
Informatics Forum, 10 Crichton Street, Edinburgh EH8 9AB, UK

Jon Oberlander (jon@inf.ed.ac.uk)
School of Informatics, University of Edinburgh
Informatics Forum, 10 Crichton Street, Edinburgh EH8 9AB, UK

Simon Kirby (simon@ling.ed.ac.uk)
Language Evolution and Computation Research Unit, University of Edinburgh
Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD, UK

Abstract

Human languages include vast numbers of learned, arbitrary signal-meaning mappings but also many complex signal-meaning mappings that are systematically related to each other (i.e. *not* arbitrary). Although arbitrariness and systematicity are clearly related, the development of the two in communication systems has been explored independently. We present an experiment in which participants invent signs from scratch to refer to a set of real concepts that share semantic features. Through interaction, the systematic re-use of arbitrary elements emerges.

Keywords: arbitrariness; systematicity; signs; language evolution; emergent communication

Introduction

Two of language's most fascinating properties, arbitrariness and systematicity, characterize the nature of the mappings between signals and meanings. A sign is *arbitrary* when there is no inherent relationship between the signal and its meaning. For example, the sounds in the word "house" have nothing to do with what the word means. In contrast, some subsets of signs in a language are *systematic*, in that signals for similar meanings share an element. The referring expressions "big house", "red house", "big apple", and "red apple" are an example. In language, words are often arbitrary while multi-word phrases are systematic. How does this property, the systematic re-use of arbitrary elements, emerge in communication systems?

Recent experimental work has shown that people are able to successfully communicate in the absence of conventional communication systems, often by creating novel signs. (de Ruiter et al., 2007; Galantucci, 2005; Garrod et al., 2007; Healey et al., 2002; Scott-Phillips, 2009). The first signs people produce in these situations are often not arbitrary, but rather iconic or motivated in some other way. (Galantucci, 2005; Garrod et al., 2007)

Psycholinguistic work has demonstrated how referring expressions can change during dialogue. (Garrod & Doherty, 1994; Pickering & Garrod, 2004). In particular, conversational partners collaborate to establish definite references, and allowing their referring expressions to shorten. (Clark & Wilkes-Gibbs, 1986). This simplification causes iconic signs in novel communication systems to become more arbitrary. (Garrod et al., 2007)

Kirby (2001) demonstrated how, given a set of arbitrary signs, systematicity might evolve. Simple artificial agents learn sets of signs and detect chance regularities in them (e.g., that the words for two red items both contain the syllable "ka"). Over many generations of agents producing signals for new meanings (meanings they didn't learn signals for) according to the regularities they observed, a set of signs can become systematic. Kirby et al. (2008) confirmed the result in human experimental participants.

Taking these two lines of research together, we have one route to the systematic re-use of arbitrary elements: people generate signs that are non-arbitrary, those signs become arbitrary as they simplify, by chance there are a few signal-meaning regularities, generations of people propagate these regularities, and the language becomes systematic. It's this longer history of a communication system, from the birth of the first sign to a set of signs which systematically re-uses arbitrary elements, that the current work aims to explore.

Goldin-Meadow et al. (1995)'s study on the emergence of systematicity in homesign (gestures created by deaf, non-signing children for use with their caretakers) covers this range. They found that, in the early stages of the homesign systems, a particular value of a particular gesture component (such as a 1" distance between the thumb and index finger) was used in gestures for just one object. In the later stages, the homesigners apparently collapsed some distinctions between objects and applied some values of gesture components to more than one object, increasing the systematicity of his or her set of gestures. This work shows that systematicity doesn't require complete arbitrariness – the recurrences between signal components and meaning components weren't chance. Unfortunately, we cannot know whether homesigners systems would have been systematic from the earliest stages, given similar-enough objects.

Here we present an exploration of the emergence of the systematic re-use of arbitrary elements in one controlled experiment. In this way, we can probe the relationship between systematicity and arbitrariness as communication systems develop.

Experiment

Methods

Participants 32 University of Edinburgh students participated in exchange for £12. All were native British English speakers. Participants who played together didn't know each other.

Apparatus Partners were seated in separate soundproof booths with computers. The experiment was run using the Pigeon software (Healey et al., 2002), which presented the item to draw each trial and provided a shared online whiteboard. Participants guessed and corrected their partners' guesses in an MSN Messenger chat window.

Stimuli The items were chosen to share salient semantic features; each item can be thought of as one of five entity types (such as person or building) that relates to one of ten themes (such as education or agriculture). There were 26 core items, appearing with different frequencies. These are shown in Figure 1. Additionally, there were 14 filler items, occurring just once per game, intended to prevent participants from assuming that their set of items was closed. The items occurred in random order. Participants knew nothing about the items in advance. In particular, were never exposed to a list of the items.

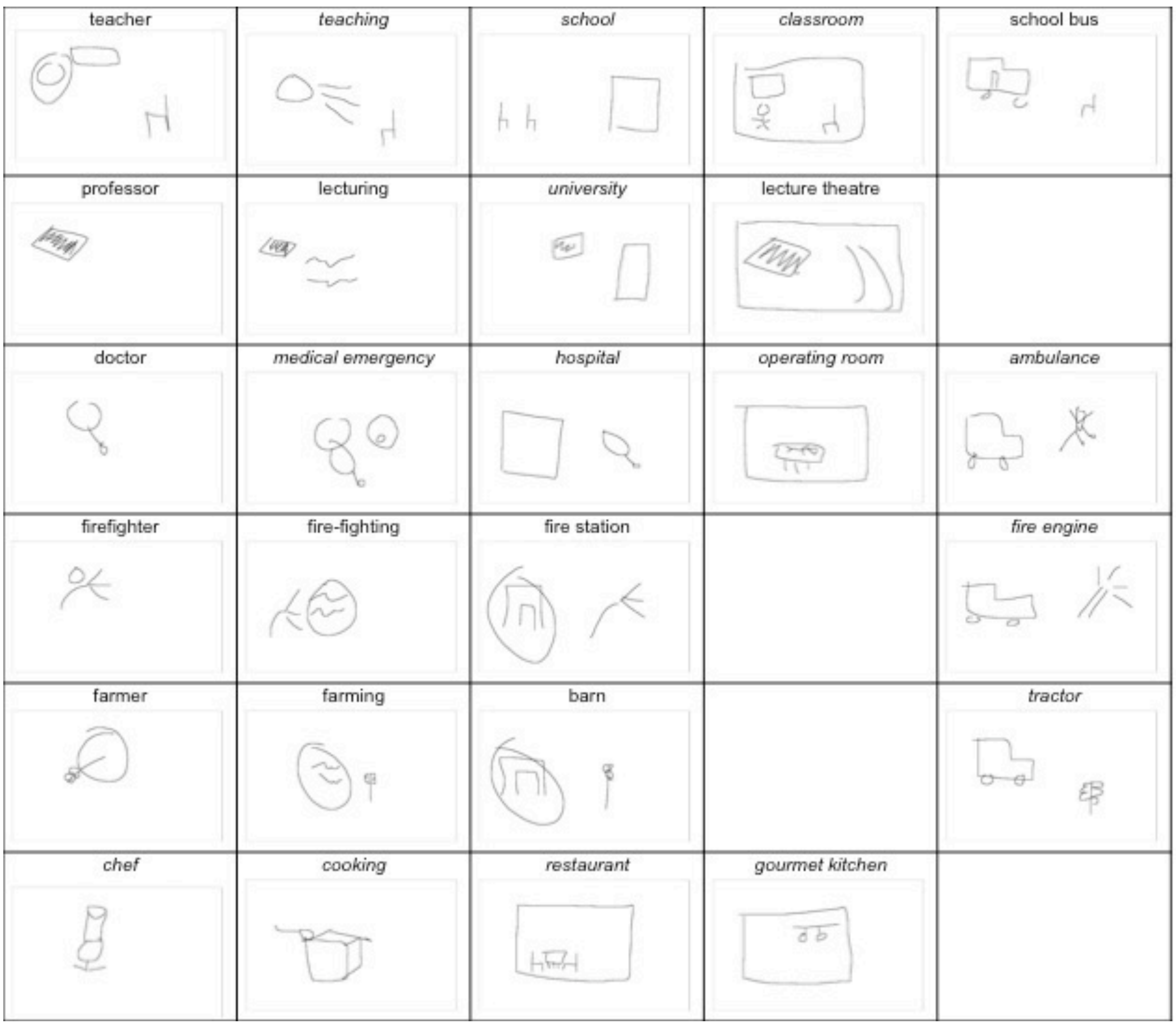


Figure 1. One set of signs that emerged from the experiment. Each sign is the last occurrence of that item in the game. Signs are arranged according to the semantic features of the items, not by chronological order of the trials. Italics distinguish which participant was the Drawer that trial. The set is highly systematic, in that signs in many of the rows and columns share an element. Also notice how arbitrary the elements have become.

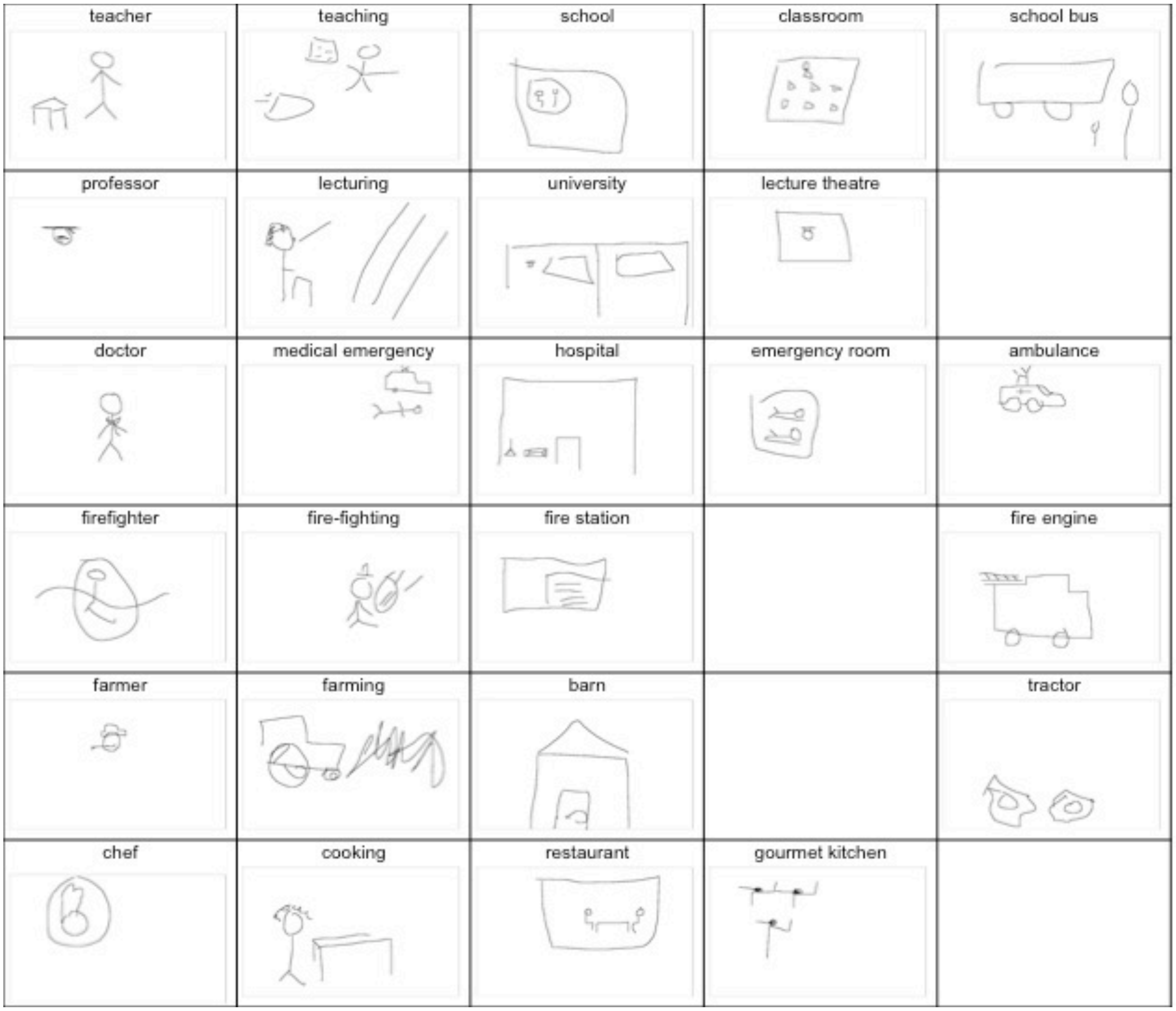


Figure 2. One Mixed Last set of signs. Each drawing in a row or column is the last from a different game of the experiment (i.e. from different pairs of players). Notice how little systematicity the set has.

Game A team was allowed just one guess per trial. A team won 1 point for every correct guess but lost 1 point for any incorrect guess or drawing that included a symbol or convention. The goal was to win as many points as possible in the two hours of play. Participants from the three top-scoring teams were entered into a prize draw for an additional £20.

Procedure Each trial, one participant was the Drawer and other was the Guesser. The Drawer saw an item (such as `professor`) on his screen and was allowed to draw immediately. The Drawer drew with a mouse, had only black ink, and could not erase anything. The Guesser saw everything the Drawer drew immediately, on her screen. The Guesser did not see the Drawer's mouse movements when he was not drawing, and could not draw herself. When she was ready, the Guesser guessed by typing into a chat window. The Drawer stopped drawing immediately and either confirmed or corrected the guess in the chat window. Players advanced themselves to the next trial. Every six trials, the participants switched Drawer and Guesser roles. The participants played for two hours.

Results: Systematicity

Figure 1 shows one of the systems that emerged from this game. Notice how systematic it is: the drawings in many of the rows and columns share an element. For example, the drawings for items relating to university education (in the second row) each have a filled-in diamond. As another example, four of the drawings for activities/situations (in the second column) have rows of squiggly lines.

To enable analysis of systematicity, each set of drawings was printed on a page in a table, organized so that rows and columns contain drawings for similar items (as in Figure 1). A single coder examined each row and each column for any element shared among two or more drawings. If there was a shared element, the coder marked which of the drawings in that row or column included it.

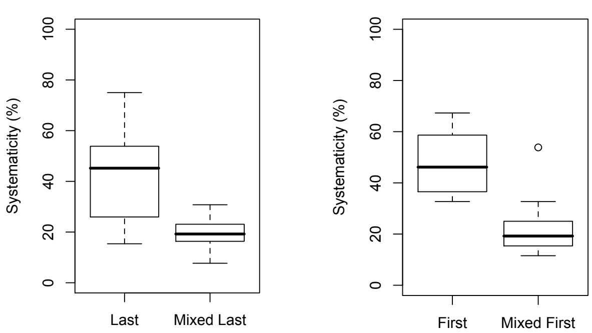
There are 26 drawings and each drawing is inspected twice – once as a member of its row and once as a member of its column. Thus, each set of drawings can receive a total score of 52. The total score divided by 52 is our systematicity score (a percentage).

For each of the 12 games from the experiment, we coded the set of First drawings (the first drawing of each item from that pair of participants) and the set of Last drawings (the last drawing of each item). To put their systematicity scores in context, we constructed 12 sets each of two kinds of comparison sets: Mixed First and Mixed Last. The Mixed First (or Last) sets were each composed of the First (or Last) drawings from *different* games of the experiment (i.e. from different pairs of players). For each Mixed set, for each item (e.g., `teacher`), we choose at random which of the games the drawing would be from, with the restriction that the drawings in each row and each column would be from different games. Figure 2 shows one Mixed Last set.

The coder marked these 48 sets in random order and blind.

A different coder marked three randomly chosen sets of each category independently. Her scores were strongly correlated with those of the original coder (Spearman's $\rho = 0.82$, $p = 0.001$).

Figure 3A shows the mean systematicity for the Last and Mixed Last sets of signs. ($M_{Last} = 42.79$, $SD = 18.95$; $M_{Mixed\ Last} = 19.39$, $SD = 6.32$) Last sets of signs are more systematic than Mixed Last sets of signs. That is, signs drawn at the end of the games re-use elements more than can be attributed to the tendency across pairs of players to draw these items the same way (roughly, iconicity) – they are truly systematic. A Mann–Whitney U Test confirmed this ($p < 0.005$).



A **B**

Figure 3. Mean systematicity (%) and confidence intervals (confidence level = 95%) for First, Mixed First, Last, and Mixed Last sets of signs. Last sets are more systematic than Mixed Last. First sets are more systematic than Mixed First.

How did Last sets of signs get to be systematic? It turns out that First sets of signs are also truly systematic. Figure 2B shows the mean systematicity for First and Mixed First sets of signs. ($M_{First} = 47.76$, $SD = 11.85$; $M_{Mixed\ First} = 22.28$, $SD = 11.67$) A Mann-Whitney U Test confirmed that the First sets of signs have significantly higher systematicity scores than the Mixed First sets of signs ($p < 0.001$). Further, as Figure 4 illustrates, there's a strong correlation between the First and Last systematicity of the sets. (Spearman's $\rho = 0.62$, $p < 0.05$).

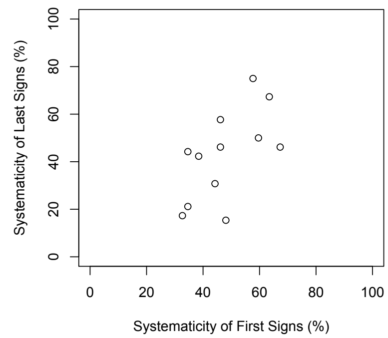


Figure 4. Scatterplot of Last against First Systematicity (%). The ranking of the First sets of signs by systematicity correlates with that of the corresponding Last sets of signs.

Results: Arbitrariness

To measure the arbitrariness of the signs produced in the experiment, we followed Fay et al. (2008) and had new participants guess what they meant. 12 University of Edinburgh students, all native British English speakers, participated in exchange for one chance in a £25 prize draw for each correct guess. The experiment was run online, and lasted approximately 15 minutes. Participants learned about the original game, and that the drawings they'd see would be from different games and different points in the games, in random order. Each trial, a participant saw a screenshot of the whiteboard at the end of the trial in the original game. He guessed the meaning of it by clicking on one of 26 buttons, one for each possible item. Each participant was presented with the First drawings of each core item from one randomly-assigned original game and the Last drawings of each core item from a different randomly-assigned original game.

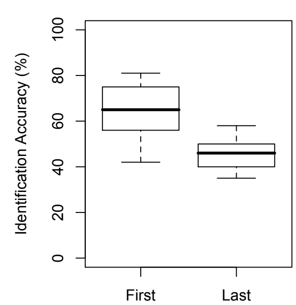


Figure 5. Mean identification accuracy (%) and confidence intervals (confidence level = 95%) for First and Last sets of signs. First sets of signs are more accurately identified than Last.

Figure 5 shows the mean identification rates (as proportions correct) for First and Last sets of signs. (M_{First} = 64.08, SD = 12.37; M_{Last} = 45.42, SD = 6.86) First sets of signs were more accurately identified than Last sets of signs. A Mann–Whitney U Test confirmed this ($p < 0.001$). This suggests that the signs became more arbitrary over the course of the games.

Discussion

We've presented an experiment in which the systematic re-use of arbitrary elements emerges. Last sets of signs are systematic, and becoming more arbitrary.

While previous work has explored the "evolution" of systematicity, this experiment has shown systematicity in the very first signs people use with each other. It appears to simply emerge, without explicit design on the part of the participants, as a natural part of dialogue.

Where does this initial systematicity come from? One might expect that the first time a player draws a certain item (say, `school bus`) with his partner, he draws it no differently than if he were drawing with a new partner. But if this were the case, the First sets would have been no more systematic than the Mixed First sets. Instead, when drawing items for the first time, players seem to have referenced previous drawings of related items. Consider Figure 6, in which one pair's first drawing of `school bus`, which occurred *after* another primary education item (`teacher`) had been drawn, is contrasted with three pairs' first drawings of `school bus`, each of which occurred *before* any other primary education item had been drawn. The former drawing for `school bus` includes elements found in the previous drawing of `teacher`, viz. the chalkboard and two children - elements not found in the other pairs' first drawings for `school bus`.

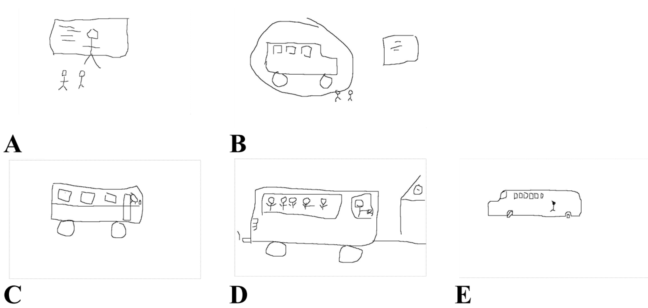


Figure 5. One pair's first drawings for `teacher` (A) and then `school bus` (B), contrasted with three other pairs' first drawings for `school bus` (C – E).

Thus, the systematicity results presented here apply not just to iconic reference to tangible objects, but to communication in general.

A common (albeit often implicit) assumption in the literature is that a novel communication system will first become arbitrary and then develop systematicity. For example, Garrod et al. (2007) say they offer an account of the "evolution of sets of icons into sets of symbols, and of sets of symbols into symbol systems." In contrast, the current work suggests that proper systematicity need not wait for arbitrariness.

Similarly, the current work shows that, as sets of signs become more arbitrary, they don't necessarily become less systematic. Structure can be retained while the elements become arbitrary. Garrod et al. (2007) suggested this, but didn't explore systematicity directly.

We've presented a paradigm that allows one to explore arbitrariness and systematicity in one experiment. Future work should explore the many issues surrounding the interaction of the two properties, as well as the transmission to others of communication systems which make systematic re-use of arbitrary elements.

Acknowledgments

We are particularly grateful to Nic Fay for formative discussions about the paradigm and data to play with, Nik Swoboda for help with and modifications to Pigeon, and the Language Evolution and Computation Research Unit for feedback at every stage and hours of piloting. This work was supported by the European Union Integrated Project JAST (Joint Action Science and Technology), grant FP6-IST2-003747.

References

Clark, H. H. & Wilkes-Gibb, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.

De Ruiter, J.P., Noordzij, M., Newman-Norlund, S., Hagoort, P., & Toni, I. (2008). On the origin of intentions. *Attention and Performance XXII: Sensorimotor foundation of higher cognition* (pp. 593–610). Oxford: Oxford University Press.

Fay N., Garrod S., Roberts L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509): 3553-61.

Galantucci, B. (2005). An Experimental Study of the Emergence of Human Communication Systems. *Cognitive Science: A Multidisciplinary Journal*, 29(5), 737-767.

Garrod, S., & Doherty, G. (1994). Conversation, Coordination and Convention - an Empirical-Investigation of How Groups Establish Linguistic Conventions. *Cognition*, 53(3), 181-215.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science*, 31(6), 961-987.

Goldin-Meadow, S. (1995). The resilience of combinatorial structure at the word level: morphology in self-styled gesture systems. *Cognition*, 56(3), 195-262.

Healey, G. T., Swoboda, N., & King, J. (2002). A tool for performing and analysing experiments on graphical communication. *People and computers XVI: Proceedings* *of HCI2002: The 16th British HCI Group Annual Conference* (pp. 55-68). London: Springer-Verlag.

Healey, P. G. T., Garrod, S., Fay, N., Lee, J., & Oberlander, J. (2002). Interactional context in graphical communication. *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 441–446). Mahwah, NJ: Laurence Erlbaum Associates, Inc.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, 5, 102–110.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS*, 105(31), 10681-10686.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-226.

Scott-Phillips, T. (2009) *The Social Evolution of Pragmatic Behaviour.* PhD Thesis, University of Edinburgh.

Systematicity and arbitrariness in novel communication systems

Carrie Ann Theisen
School of Informatics; University of Edinburgh
carrieanntheisen@gmail.com
10 Crichton Street; Edinburgh EH8 9AB; UK

Jon Oberlander
School of Informatics; University of Edinburgh
jon@inf.ed.ac.uk
10 Crichton Street; Edinburgh EH8 9AB; UK

Simon Kirby
School of Philosophy, Psychology and Language Sciences; University of Edinburgh
simon@ling.ed.ac.uk
3 Charles Street; Edinburgh EH8 9AD; UK

Arbitrariness and systematicity are two of language's most fascinating properties. Although both are characterizations of the mappings between signals and meanings, their emergence and evolution in communication systems has generally been explored independently. We present an experiment in which both arbitrariness and systematicity are probed. Participants invent signs from scratch to refer to a set of items that share salient semantic features. Through interaction, the systematic re-use of arbitrary signal elements emerges.

Keywords: arbitrariness; systematicity; signs; language evolution; emergent communication

Biographical note: Carrie Theisen is currently a PhD student at the University of Edinburgh's School of Informatics, and affiliated with the Language Evolution and Computation group there. She earned a BA and a BS from the University of Pittsburgh with majors in (Cognitive) Psychology, Linguistics, and Philosophy and several distinctions. As a Marshall Scholar, she went on to earn an MSc in Informatics (Natural Language and Language Engineering concentration) from the University of Edinburgh. Her research experiences have focused on how human communication (especially reference) works.

Introduction

Two of language's most fascinating properties, arbitrariness and systematicity, are characterizations of the mappings between signals and meanings. Think of a language as a set of signs, or signal-meaning mappings, at different levels, e.g., morphemes, words, phrases, etc. A sign is *arbitrary* when there is no inherent relationship between its signal and its meaning. For example, the sounds in the word "house" have nothing to do with what the word means. Rather, the signal is attached to the meaning by convention. Following Peirce (1955), we can contrast arbitrary signs with icons (in which the signal resembles its meaning, as a sketch of an item does) and indices (in which the signal is directly connected to its meaning, e.g., physically or causally, as smoke signals fire because fire causes smoke).

We can also consider how *sets* of signals relate to their meanings. Some sets of signs in a language are *systematic*, in that signals for similar meanings share an element. For example, the phrases “small house”, “expensive house”, and “red house” all include “house” and all refer to buildings in which people live. In contrast, idioms make a language less systematic. For example, “bought the farm” means something completely different from what similar signals (such as “bought the house” or “saw the farm”) mean. Likewise, signals with meanings similar to “bought the farm” (such as “died”) bear no similarity to it.

The “small house”, “expensive house”, and “red house” examples above illustrate a particular way arbitrariness and systematicity work together in language: we often find that words are arbitrary while multi-word phrases are systematic. How does this property, the systematic re-use of arbitrary elements, emerge in a communication system?

Previous Work

Recent work has pioneered the experimental study of how communication systems emerge and evolve. A fundamental result of this work is that people can find ways to successfully communicate in the absence of conventional communication systems, often by creating novel signs. For example, de Ruiter et al. (2007) had subjects play a video game in which they had to cooperate in order to move their tokens into the correct orientation and location. Similarly, partners in Scott-Phillips et al. (2007)’s game were to move to rooms of the same color, but could not see the colors of each other’s environments. In both cases, although players had only the motion of their own tokens with which to communicate, most were eventually successful.

The first signs people produce in these situations are usually iconic or motivated in some other way, i.e. *not* arbitrary[1]. For example, the task of Galantucci (2005)'s subjects was to bring their agent into the same room as their partner. Each of the rooms was marked with an icon, but partners could communicate using only a novel graphical channel that transformed their signals in an obscure way. After a number of attempts, some subjects established signs that allowed them to win the games. The signals for the rooms varied widely across the games but many had inherent connections to their meanings: representing the room's location in the environment or the shape or the number of vertices of the room's icon. People can find ways to graphically represent even music. Healey et al. (2007) presented a Music Drawing Task, in which subjects refer to short pieces of unfamiliar music with their partners. They can draw on a virtual whiteboard, but may not write letters or numbers. Most of the signals they produce to solve the task make use of one of two strategies. Some drawings consist of recognizable objects, figures, or scenes – these are iconic drawings of associations one might make to the pieces of music, e.g. a racecar for a piece of music with a fast tempo. Other drawings represent the structure of the pieces, e.g. drawing a curve with peaks and valleys that match the high and low notes in the piece. Clearly, people can exploit a wide range of connections between signals and meanings to get communication systems off the ground.

If the first signs people create are not arbitrary, how do they become arbitrary? Some of this new experimental work has subjects refer to the same item several times, giving us the opportunity to examine how a signal for a given meaning evolves. The Concept Drawing Task (Garrod et al., 2007; Healey et al., 2002) is particularly

---

[1] Scott-Phillips et al. (2007) offer an interesting exception.

promising. Each round, one of the subjects (the "Drawer") has an ordered list of concepts and the other (the "Chooser") has an unordered list of the same concepts. The Drawer takes each of the concepts in turn and produces a sketch for it on virtual whiteboard they share, so that the Chooser can identify the concept. Players can draw whatever they like but may not write anything, such as letters or numbers. Pairs play for a number of blocks, so that each concept is drawn and identified several times over the course of the game. Among other measures of how the signals for the items changed with interaction, Garrod et al. (2007) found that graphical complexity (roughly, the amount of virtual ink used) decreases. They argue that this complexity is a measure of iconicity – all else equal, the less information there is in the signal, the less the signal can resemble its meaning – thus showing that signs become more arbitrary with interaction.

We now see how arbitrariness can emerge in a novel communication system; what about systematicity? These above studies don't measure the degree to which signals for similar meanings share an element.[2] Some provide examples of novel sets of signs that appear to exhibit systematicity (Galantucci, 2005; de Ruiter et al., 2007; Healey et al., 2007), but the recurring elements appear to be iconic.

Garrod et al. (2007) provide an example of the systematic re-use of arbitrary elements: one pair's successive drawings for "art gallery". Each contains three elements: an iconic drawing of a building, something like a "less than" symbol, and an iconic drawing of a painting. They note that the building element became associated with all and only the institutions in the set of concepts, and describe how the building and painting elements simplified to become more arbitrary. As they don't actually

[2] Healey et al. (2007) offer a measure of a sense of "systematicity" (viz. the use of a common frame of reference when producing signals) that is more general than ours here.

measure systematicity, it's left open how common this systematicity is across pairs, and how early this systematicity emerges with respect to the evolution of arbitrariness. For example, did the building element occur in drawings for institutions other than "art gallery" (say, "parliament") while it was still very iconic? Or did the pair change its drawings for the other institutions later in the game?

Kirby (2001) simulates how systematicity can evolve. Simple artificial agents learn sets of arbitrary signs. The set of meanings and the set of signals are both highly structured. The agents detect chance regularities in them (e.g., that the words for two red items both contain the syllable "ka"). Over many generations of agents producing signals for new meanings (meanings they didn't learn signals for) according to the regularities they detected, sets of signs become more systematic. Kirby et al. (2008) confirmed the result in human experimental participants. This work takes a set of arbitrary signs as its starting point, but we saw above that the first signs people create are not arbitrary. They can become arbitrary by simplifying, but it's not clear that the signals would still share elements once they did. (Of course, the signals need to share elements if the language is to become systematic.) Further, if the first signs that people create can be systematic (using iconic elements), it seems odd that they would become completely arbitrary (i.e. lose all this systematicity) and then become systematic again.

To summarize, recent experimental work has demonstrated how novel signs can emerge and then evolve to become arbitrary, but does not measure systematicity. Kirby (2001) and Kirby et al. (2008), on the other hand, propose a mechanism for and measure the evolution of systematicity, but assume an artificial stage in the evolution of novel sign systems. Thus, it's not clear how the systematic re-use of arbitrary elements emerges in novel communication systems. The current work aims to fill this gap, by presenting an experiment in which both arbitrariness and systematicity can be probed.

Current Approach

We present an experiment that combines aspects from several of the above approaches to explore the emergence of the systematic re-use of arbitrary elements.

Two participants are seated in separate, soundproof computer booths. On each of their screens is a (blank) virtual whiteboard and a chat window, both of which they share. Each trial, one is the Drawer and the other the Guesser. The Drawer sees an item (such as 'teacher") on his screen. His task is draw on the whiteboard with a mouse in order to get his partner to guess the item. Neither the Drawer nor the Guesser has seen a list of the items that will appear in the game – he knows nothing about what other items they will eventually communicate to each other, and she must guess off the top of her head. The Drawer may draw anything he likes on the whiteboard, except he may not use any pre-established signs (e.g., letters, arrows, a heart to represent love, etc.). The Guesser sees everything the Drawer is drawing. She has just one guess, but is motivated to guess quickly in order to score more points in the game. When she's ready, she types her guess into the chat window. He stops drawing and types the item into the chat window (whether her guess was correct or incorrect). Nothing but the guess and the actual item are allowed in the chat window. The participants proceed to the next trial, with a new item chosen at random. The participants switch Drawer and Guesser roles every 6 trials, so that both have a hand in creating the signs. They play for 2 hours, during which time the same 26 items appear over and over again. This design allows us to examine what signs the participants create for each item, as well as how they evolve over the course of the game.

The task is adaptation of the Concept Drawing Task presented in Healey et al.

(2002) and Garrod et al. (2007). The most significant difference is in the set of items. Following Kirby (2001) and Kirby et al. (2008), the items share salient semantic features with each other. This way, we can measure the degree to which the drawings for similar items (items that share a semantic feature) share an element, i.e. systematicity.

We hypothesize that an increase in arbitrariness need not entail a decrease in systematicity. In particular, we expect that the first signs people create will exhibit systematicity, but will be non-arbitrary. With interaction, the signal elements will become more arbitrary, but the systematicity of the set will be retained.

Experiment

Participants

24 University of Edinburgh students participated in exchange for £12. All were native British English speakers. None had a background in Linguistics (in case an awareness of how language works would influence the signs they created). Participants who played together didn't know each other.

Apparatus

Partners were seated in separate soundproof booths with computers. The experiment was run using the Pigeon software (Healey, Swoboda, & King, 2002), which presented the item to draw each trial and provided a shared online whiteboard. Participants guessed and corrected/confirmed their partners' guesses in a chat window.

Stimuli

The items are listed in Figure 1. They were chosen to share salient semantic features; each item can be thought of as one of five entity types (such as person or building) that relates to one of ten themes (such as education or agriculture).

The 26 core items (boxed in Figure 1) appeared with different frequencies to approximate real life, in which people communicate about some things more often than others. Pilot experiments determined that participants were likely to complete at least 126 trials during the allotted time for the game, so the frequencies of the items was determined for blocks of 126 trials. Every 126 trials, each of the four items in the innermost box in Figure 1 appears 8 times, each item in the next box out appears 5 times, and each item in the next box out appears 3 times. Additionally, there were filler items (unboxed in Figure 1), meant to stop participants from assuming that their set of items was closed. Every 126 trials, six of the unboxed items appear. Each of these unboxed items appears at most once per game.

The items were randomized within each block of 126 trials.

teacher	school	teaching	school bus	classroom
firefighter	fire station	fire-fighting	fire engine	
professor	university	lecturing		lecture theatre
doctor	hospital	medical emergency	ambulance	emergency room
chef	restaurant	cooking		gourmet kitchen
farmer	barn	farming	tractor	
soldier	barracks	war	tank	
prisoner	jail	crime	police car	
chemist	pharmacy	prescription		
dentist	dental practice	root canal		

Figure 1. The items. Boxes indicate their relative frequencies, where the items in the innermost box appear most frequently.

Rules of the Game

The experiment is meant to tell us something about how people communicate in the absence of conventions, but people have drawn and seen drawings before. So we prohibited the use of symbols and drawing conventions in the game: letters, numbers, punctuation (?, !, etc.), mathematical signs (+, =, <, etc.), pointing with arrows, universal signs (the cross for first aid, a skull and cross-bones for danger, etc.), drawing conventions (a heart to represent love, + over the eyes to show that someone is dead, etc.).

A pair was allowed just one guess per trial. After the Guesser guessed, the Drawer typed the actual item, so that both knew whether she was correct. A team won 1 point for every correct guess but lost 1 point for any incorrect guess or drawing that included a symbol or convention. We penalized incorrect guesses to discourage too-early guesses. The goal was to win as many points as possible in the 2 hours of play. This was to encourage participants to get through as many trials as possible so that we had many drawings to analyze. It also introduced a pressure to minimize production effort. Participants from the three top-scoring teams were entered into a prize draw for an additional £20.

Procedure

Before the experiment started, participants read instructions. They did not see a list of the items they were about to communicate, nor know anything about the items.

Each trial, one participant was the Drawer and other was the Guesser. The Drawer saw an item (such as "professor") on his screen and was allowed to draw immediately. The Drawer drew with a mouse, had only black ink, and could not erase anything. These restrictions were intended to limit the amount of information that the Drawer could encode, hopefully allowing the participants to get through more trials. The Guesser saw everything the Drawer drew immediately, on her screen. The Guesser did not see the Drawer's mouse movements when he was not drawing, and could not draw herself.

When she was ready, the Guesser guessed by typing into a chat window. She was only allowed one guess. The Drawer stopped drawing immediately and typed the item into the chat window, whether or not the guess was correct. Nothing but the guess and the exact item were allowed in the chat window. The chat window was sized to show only the previous three guesses or items, and participants were not allowed to scroll up to see earlier items. Once the Guesser learned what the item was, she hit a Next Item button. Both participants hit an OK button to start the next trial.

Every six trials, the participants switched Drawer and Guesser roles. The participants played for two hours.

The instructions to the participants are included as Appendix A.

Results: An Example Sign System

Figure 2 shows one of the sets of signs to emerge from this game. Each image shows what was drawn on the whiteboard for that item the last time it occurred in the game, regardless of which participant was the Drawer that trial. Items written in italics were drawn by one participant, while items in the normal typeface were drawn by the other. The signs are arranged according to the semantic features of the items, not by the chronological order of the trials. Items in the same row and items in the same column share a semantic feature. For example, the first column contains this pair's drawings for various people.

[INSERT FIGURE 2 ABOUT HERE.]

The first thing to notice is how systematic the set is: the drawings in many of the rows and columns share an element. For example, the drawings for items relating to university education (in the second row) each have a filled-in diamond. As another example, four of the drawings for activities/situations (in the second column) have a row of squiggly lines.

Also notice how arbitrary some of the elements have become. It is probably not transparent to the reader that the filled-in diamond recurring in drawings for university education items evolved from a drawing of a professor wearing a mortarboard.

Results: Systematicity

We measured the systematicity of each pair's initial set of signs (the drawings produced the first time each of the items appeared in their game) and final set of signs (the drawings produced the last time each of the items appeared in their game), to determine whether systematicity emerged and, if so, when. We hypothesize that both the initial sets of signs and the final sets of signs will exhibit systematicity and, further, that degree of systematicity will be retained from the initial sets of signs to the final.

Measuring Systematicity

Kirby (2001) and Kirby et al. (2008) measured systematicity as a Pairwise Distance Correlation: the correlation of the similarities between signals with the similarities between their meanings is calculated, and then compared with the degree of correlation one would expect if the signals had been randomly assigned to meanings. RegMap (Tamariz & Smith, 2008) measures the degree of regularity in the mappings between signals and meanings without such similarity measures. Instead, each possible mapping between semantic feature and signal element is considered, and the number of examples of the mapping is compared to the number of counterexamples. The signals produced in our experiment are not necessarily composed of discrete elements, though, so both of these methods would require a great many human judgments.

We adapted these measures of systematicity to our data. In a nutshell, we consider each semantic feature in turn, look at the drawings for items that share that feature, count the number of those drawings that share an element, and sum these counts to get a measure of systematicity for the whole set of drawings. We then compare this to the amount of systematicity we'd expect by chance.

Coding for Systematicity

One of the experimenters coded each set for systematicity. Each set of drawings was printed on a page in a table, organized so that rows and columns contained drawings for similar items. The coder examined each row and each column for any element shared among two or more drawings. If there was a shared element, the coder marked which of the drawings in that row or column included it. The coding instructions are included as Appendix B.

There are 26 drawings and each drawing is inspected twice – once as a member of its row and once as a member of its column. Thus, each set of drawings can receive a total score of 52. The total score divided by 52 is our systematicity score.

For each of the 12 games from the experiment, we coded the Initial sets of drawings (the first drawing of each item from that pair of participants) and the Final sets of drawings (the last drawing of each item). To obtain a measure of baseline systematicity, we constructed 12 sets each of two kinds of control sets: Mixed Initial and Mixed Final. The Mixed Initial sets were each composed of the initial drawings from *different* games of the experiment (i.e. from different pairs of players). For each Mixed set, for each item (e.g., "teacher"), we choose at random which of the games the drawing would be from, with the restriction that the drawings in each row and each column would be from different games. Thus we have four categories of sets of signs, each including 12 sets of signs: Initial, Final, Mixed Initial, and Mixed Final.

The coder marked these 48 sets in random order and blind, i.e. she didn't know which set she was coding.

Reliability

To check whether the experimenter's coding was biased, a different coder – who had no role at all in the experiment – marked three randomly-chosen sets of each category (12 sets in total) independently. Her systematicity scores were strongly correlated with those of the original coder (Spearman's $\rho = 0.82$, $p = 0.001$). (The sets of systematicity scores failed standard tests for normality, and so non-parametric statistics were used here and in subsequent systematicity analyses.)

Results

Figure 3 shows the mean systematicity for each of the four categories of sets of signs[3]. ($M_{Final} = 42.79$, SD = 18.95; $M_{Mixed\ Final} = 19.39$, SD = 6.32; $M_{Initial} = 47.76$, SD = 11.85; $M_{Mixed\ Initial} = 22.28$, SD = 11.67) The first thing to notice is that Final sets of signs are more systematic than Mixed Final sets of signs (Figure 3A). That is, signs drawn at the end of the games re-use elements more than can be attributed to the tendency across pairs of players to draw these items the same way (roughly, iconicity) – they are truly systematic. A Mann–Whitney U Test confirmed this ($p=0.002$).

[INSERT FIGURE 3 AROUND HERE.]

How did Final sets of signs get to be systematic? It turns out that Initial sets of signs are also truly systematic; see Figure 3B. A Mann-Whitney U Test confirmed that the Initial sets of signs have significantly higher systematicity scores than the Mixed Initial sets of signs ($p < 0.001$). Further, sets of signs didn't become significantly more or less systematic over time. A Wilcoxon Signed-Rank Test found no significant difference between the systematicity scores of Initial and Final sets of signs (p=

[3] One of the Mixed Initial sets of signs is an outlier, with a systematicity score more than 1.5 times the interquartile range above the third quartile for Mixed Initial sets of signs. We chose not to exclude it from analysis because there appears to be no error in creation or measurement of that set; it's simply unusually systematic for Mixed Initial sets of signs.

0.2662).

What is the relationship between a set of signs' Initial and Final systematicity then? Figure 4 is a scatterplot of each of the 12 sets' Initial systematicity score against its Final systematicity score. Clearly, there is some correlation between the Initial and Final systematicity of the sets. (Spearman's $\rho = 0.62$, $p = 0.03$)

[INSERT FIGURE 4 AROUND HERE.]

Systematicity Discussion

Sets of signs produced by partners in the same game are more systematic than signs taken from different games. This effect of interaction is present not just at the end of the games but rather right from the first signs produced.

It may seem odd that sets of signs are immediately systematic. Recall that participants have no access to the list of items they will be communicating before the game; it's not that participants are designing systematic sets of signs in advance. It should also be emphasized that players have no direct pressure for systematicity. Each pair's goal is simply to communicate accurately and quickly. In fact, there is virtually no correlation between the initial systematicity of the sets and progress in the game (i.e., how many trials were completed in the 2 hours) or final systematicity and progress. Spearman's ρ equals 0.01 ($p = 0.97$) in the former case and 0.07 ($p = 0.81$) in the latter.

One might expect that the first time one draws a certain item (say, "school bus") with his partner, he would draw it no differently than if he were drawing with a new partner. But most (at least 21 of the 26) initial drawings were produced after a related item (from either the same row or same column) had been drawn. Perhaps players are referencing these previous drawings. Figure 5A shows four pairs' first drawings of "school bus". In each of these games, "school bus" was the first primary education to be drawn, so the drawings reflect what people draw for "school bus" when uninfluenced by previous drawings of primary education items. In contrast, Figure 5B shows a fifth pair's first drawings of "teacher" and then "school bus". In that game, the drawing of "school bus" was produced *after* a drawing of another primary education item ("teacher") had been produced, and appears to re-use elements from it. In particular, note that none of the uninfluenced drawings of "school bus" include a chalkboard.

[INSERT FIGURE 5 ABOUT HERE.]

Results: Arbitrariness

As in the systematicity analyses, we examined the arbitrariness of the initial signs and final signs from each pair. We hypothesized that, as in Garrod et al. (2007), the arbitrariness of the signs had increased through interaction.

Procedure

To measure how arbitrary the signs produced in the experiment are, we followed Fay et al. (2008) and had new participants guess what they meant.

Twelve University of Edinburgh students participated in exchange for chances in a prize draw for a £25 Amazon gift voucher. All were native British English speakers. The corpus was sampled once across participants. A mixed design was used: each participant was presented with the initial drawings of each core item from one original game and the final drawings of each core item from a different original game. The participants were assigned games at random.

The experiment was run online, and lasted approximately 15 minutes. Participants read about the basics of the original communication games, and learned that the drawings they'd see would be from different games and different points in the games, in random order. Participants received an entry in the prize draw for each correct guess. Each trial, a participant saw a screenshot of the whiteboard at the end of the trial in the original game. He guessed the meaning of each image by clicking on one of 26 buttons, one for each possible item.

Results

Figure 6 shows the mean identification rates (as proportions correct) for Initial and Final sets of signs. ($M_{Initial}$ = 64.08, SD = 12.37; M_{Final} = 45.42, SD = 6.86) Initial sets of signs were more accurately identified than Final sets of signs. A Mann–Whitney U Test confirmed this ($p < 0.001$). (The arbitrariness scores were not normally distributed, so non-parametric statistics were used.) This suggests that the signs became more arbitrary (i.e. their forms related less transparently to their meanings) over the course of the games.

[INSERT FIGURE 6 AROUND HERE.]

We can also compare the arbitrariness of the drawings produced by pairs who completed more or less trials. There is a non-significant correlation between the identification accuracy of Final sets of signs and progress in the game ($\rho = -0.35$, $p = 0.26$). The correlation is negative, indicating that pairs who completed more trials in the game came to use more arbitrary drawings by the end of their games. Perhaps the more trials a pair gets through, the more arbitrary their signs become. Or maybe more arbitrary signs enable a pair of players to get through more trials (say, because they are quicker to produce).

Arbitrariness Discussion

The arbitrariness results were as predicted: signs become more arbitrary as pairs interact.

Garrod et al. (2007) argued that a sign becomes more arbitrary as information is moved from the signal itself to the history of interaction between the communication partners. If simplification is the mechanism by which signs become arbitrary, then – since the signs produced in this experiment became arbitrary over time – we expect them to have simplified as well. Preliminary observations suggest that this is indeed the case: Initial signs appear to be graphically more complex than Final signs.

Final Discussion

We have presented an experimental design in which both arbitrariness and systematicity can be probed. Further, we've demonstrated how the systematic re-use of arbitrary elements can emerge in novel sign systems. The first signs people create exhibit systematicity: signals for similar meanings share an element. As communication partners used the signs more and more with each other, the signs became more arbitrary. Yet the systematicity of the set of signs was retained.

While previous work has explored the *evolution* of systematicity in experimenter-constructed sets of arbitrary signs, our work has shown systematicity simply emerging when communication partners create their own signs. Further, the systematicity appears to be easily maintained. Could research on how communication systems *become* systematic be misguided? Perhaps we should explore instead the conditions under which initial systematicity is retained.

We posited a tendency of communication partners to reference previously used signs when creating new ones. What causes this? A Drawer might reference a previous drawing of "teacher" in his current drawing of "school bus" because he knows that his partner will understand that clue. Or the cause may be more egocentric:

perhaps the first thing a Drawer thinks of when he sees he is to communicate “school bus” is how he or his partner recently communicated “teacher”. We also suspect that *not* referencing a previous drawing of a closely-related item is somehow marked.

Previous work demonstrated that there are many different ways a signal can have some inherent connection to a meaning, i.e. there are many different ways to successfully communicate a concept for the first time. Here we see that an inherent connection between signal and meaning may be necessary only when communicating a concept completely unrelated to any you’ve communicated before. Taken together, we see how communication in a modality less well suited to iconicity, such as speech, could have emerged. Thus, the results presented here inform not just graphical communication, but communication in general.

Acknowledgments

We are grateful to the Language Evolution and Computation Group for hours of piloting and feedback at every stage of the project, Nik Swoboda for help with and substantial modifications to Pigeon, Nic Fay for formative discussions about the paradigm and data to play with, and Bruno Galantucci and an anonymous reviewer for excellent suggestions that greatly improved this paper. This work was supported by the European Union Integrated Project JAST (Joint Action Science and Technology), grant FP6-IST2-003747.

References

De Ruiter, J., Noordzij, M., Newman-Norlund, S., Hagoort, P., & Toni, I. (2007). On the origin of intentions. In P. Haggard, Y. Rossetti & M. Kawato (Eds.), *Attention & Performance XXII: Sensorimotor foundation of higher cognition* (pp. 593-610). Oxford: Oxford University Press.

Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1509), 3553-3561.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science: A Multidisciplinary Journal, 29*(5), 737-767.
Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science: A Multidisciplinary Journal, 31*(6), 961-987.
Healey, P., Garrod, S., Fay, N., Lee, J., & Oberlander, J. (2002). *Interactional context in graphical communication.* Paper presented at the Proceedings of the 24th Annual Conference of the Cognitive Science Society.
Healey, P., Swoboda, N., & King, J. (2002). A tool for performing and analysing experiments on graphical communication. *PEOPLE AND COMPUTERS*, 55-68.
Healey, P., Swoboda, N., Umata, I., & King, J. (2007). Graphical Language Games: Interactional constraints on representational form. *Cognitive Science: A Multidisciplinary Journal, 31*(2), 285-309.
Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation, 5*(2), 102-110.
Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681.
Peirce, C. (1955). *Philosophical writings of Peirce*: Courier Dover Publications.
Scott-Phillips, T., Kirby, S., & Ritchie, G. (2007). Signalling signalhood: An exploratory study into the emergence of communicative intentions. *Language, Games, and Evolution*, 77.

Appendix A: Instructions to Participants

You two are going to communicate items to each other by drawing. You'll take turns being the Drawer and the Guesser.

When you're the Drawer, you'll see an item (such as Arnold Schwarzenegger) at the top of your screen. You want your partner to guess exactly that. You may NOT use any symbols or drawing conventions, such as:

- letters
- numbers
- punctuation (?, !, etc.)
- mathematical signs (+, =, <, etc.)
- pointing with arrows

- universal signs (the cross for first aid, a skull and cross-bones for danger, etc.)
- drawing conventions (a heart to represent love, + over the eyes to show that someone is dead, etc.)

Imagine your partner is an alien who has spent time on Earth but has never drawn, written, or read anything before. If the alien wouldn't understand it, it's not allowed!

When you're the Guesser, as soon as you can, guess by typing into the chat window. You may not type anything other than your actual guess. (For example, you may NOT type *I think it's Arnold Schwarzenegger* or *Arnold Schwarzenegger?*) You may not scroll up in your chat window to remind yourself of previous guesses. Your team gets just one guess per item.

As the Drawer, you must stop drawing immediately. Type exactly what your item was, and nothing else. (For example, you may NOT type *Right!* or *You were close - it was Arnold Schwarzenegger.*)

Your team gets 1 point for every correct guess and loses 1 point for any incorrect guess or drawing that included a symbol or convention. Your team's goal is to get as many points as possible in the allotted time. Subjects from the top three teams will be entered into a prize draw for an additional 20 GBP!

The items are generated randomly. You may get the same item twice, even in a row.

Appendix B: Instructions for Coding Systematicity

You're going to see sets of drawings from a communication game. In it, pairs of people communicated items (such as "teacher" or "medical emergency") to each other by drawing on a whiteboard. Partners played with each other for a very long time, drawing the same 26 items for each other over and over again.

You're going to see sets of drawings. Each set contains a drawing for each of the 26 items (below), and represents what players were drawing at a point in time in their game.

	people	*activites / situations*	*buildings*	*rooms*	*vehicles*
primary education	teacher	teaching	school	classroom	school bus
university education	professor	lecturing	university	lecture theatre	
medical emergencies	doctor	medical emergency	hospital	emergency room	ambulance
fire-fighting	firefighter	fire-fighting	fire station		fire engine
farming	farmer	farming	barn		tractor
gourmet cooking	chef	cooking	restaurant	gourmet kitchen	

I want to know how *systematic* each set of signs is, i.e. to what extent drawings for similar items have an element in common. Notice that the drawings in each row and in each column will be for similar items. For example, the drawings in the fourth row of each set will be for items that relate to fire-fighting. You're going to look at each row and each column and tell me if any of the drawings in it have an element in common. If so, briefly describe or draw the element and mark which of the drawings have it.

Only pick out an element if you're sure that there was a special understanding between the players to draw a certain thing (or things) a certain way to mean that those items relate to fire-fighting, that those items are buildings, etc.

It's completely fine if the drawings in a row or a column have nothing at all in common. You may even see totally unsystematic *sets* of signs, in which none of the rows or columns have anything in common.

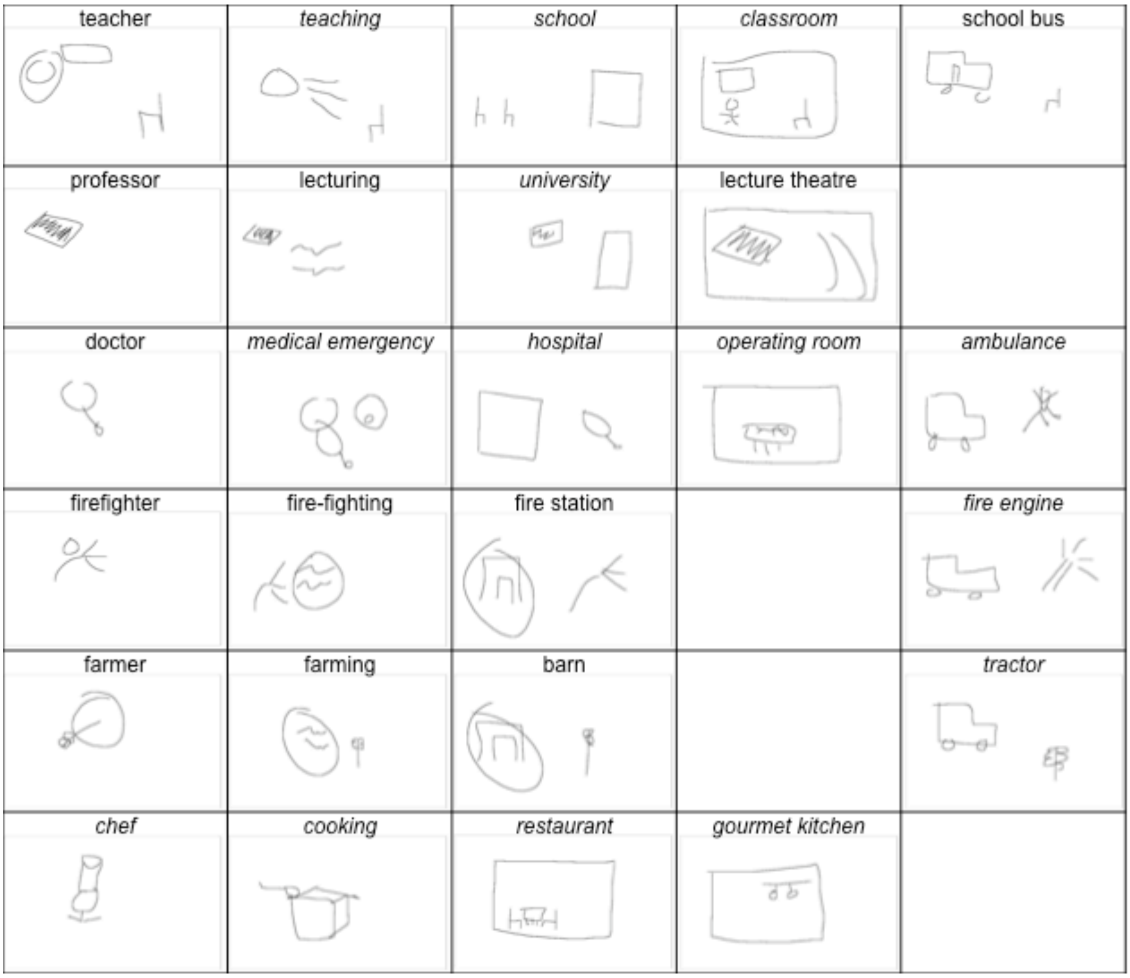
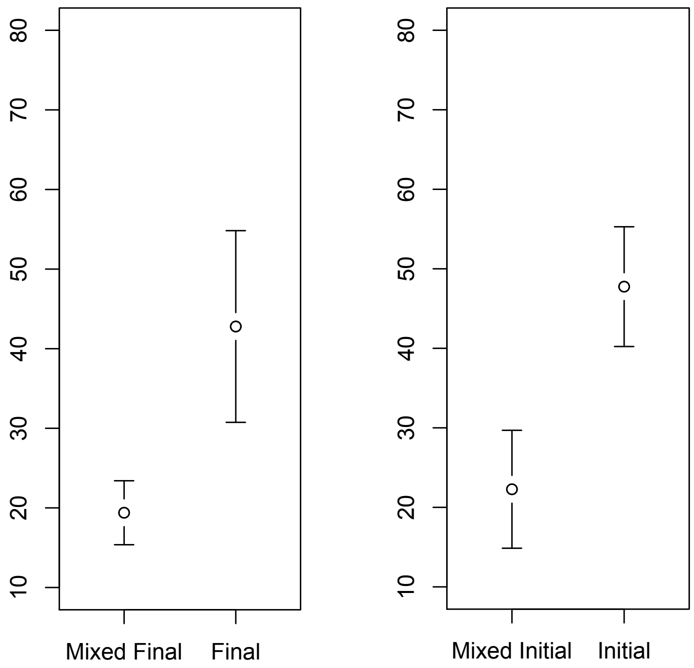


Figure 2. One set of signs to emerge in the experiment. The set is highly systematic, and the elements have become fairly arbitrary.



A B

Figure 3. Mean systematicity (%) and confidence intervals (confidence level = 95%) for Initial, Mixed Initial, Final, and Mixed Final sets of signs. Final sets are more systematic than Mixed Final and Initial sets are more systematic than Mixed Initial.

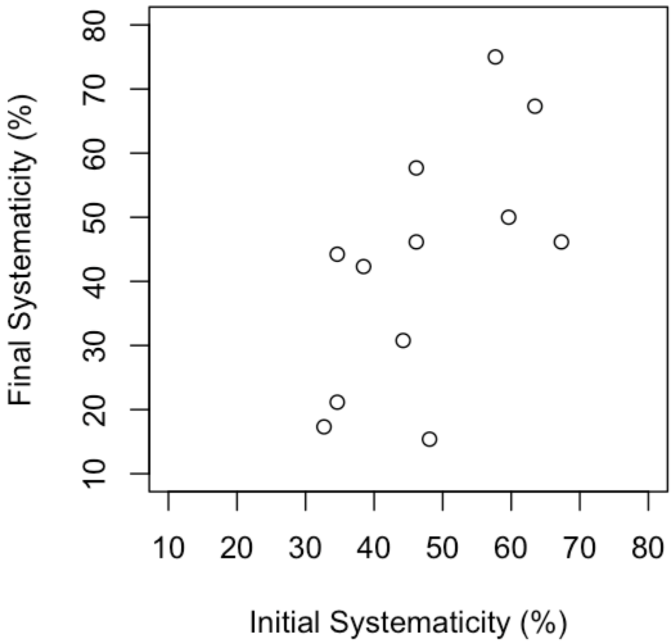
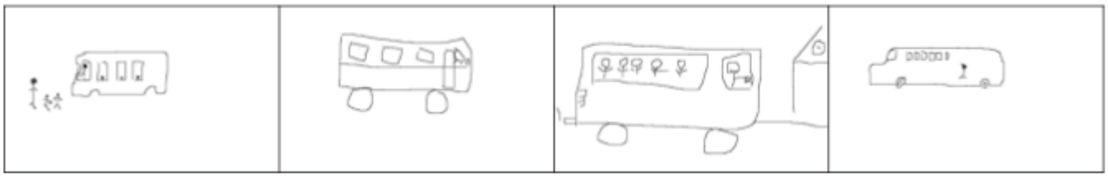
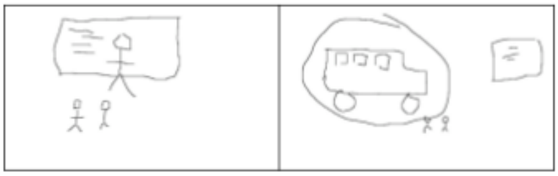


Figure 4. Scatterplot of Final against Initial Systematicity (%). There is some correlation between the ranking of the Initial sets of signs by systematicity and that of the corresponding Final sets of signs.



A



B

Figure 5. A shows the first drawings of "school bus" from four different games in which it was the first primary education item to be drawn in the game. B shows one pair's first drawings of "teacher" and "school bus". The latter was drawn after the former, and appears to re-use elements from it.

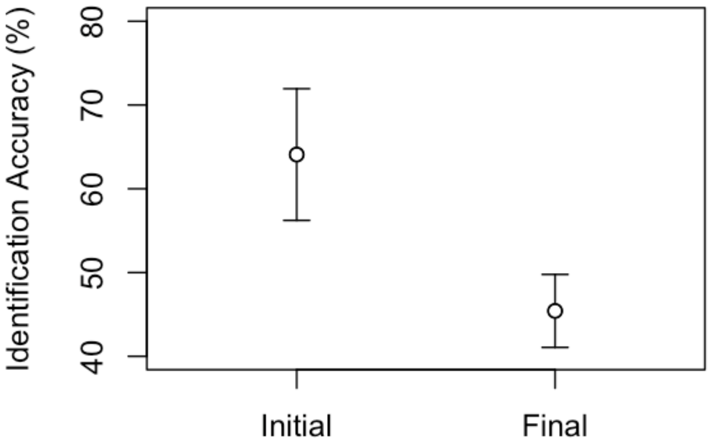


Figure 6. Mean identification accuracy (%) and confidence intervals (confidence level = 95%) for Initial and Final sets of signs. Initial sets of signs are more accurately identified than Final.